



**Ciências
ULisboa**

Tecnologia de Bases de Dados

Código: 425152

ECTS: 6

Ano Letivo: 2015/16

Carga horária: T: 2:00 h; TP: 1:30 h; OT: 2:00 h;

Departamento: Informática

Área Científica: Informática;

Objetivos da Unidade Curricular

Apresentar os conceitos e tecnologias que constituem as fundações dos sistemas de gestão de bases de dados (SGBDs), com ênfase nos SGBDs relacionais. Os SGBDs são dos sistemas mais complexos e abrangentes da Informática, sendo muitas das suas técnicas largamente empregues no desenvolvimento de muitos outros sistemas e aplicações. Ao fazer uma dissecação dos SGBDs modernos, a disciplina ilustra o seu funcionamento à luz das noções adquiridas nas disciplinas fundamentais da formação em Informática, desde a lógica e algoritmos, aos sistemas operativos e engenharia da programação.

Pré-requisitos

Sem pré-requisitos

Conteúdos

Componentes de um sistema de gestão de bases de dados. Organização física de dados e métodos de acesso. Processamento de interrogações, realização de operadores relacionais, e afinação de desempenho. Gestão de transações. Controlo de concorrência e recuperação de faltas. Segurança de informação em bases de dados relacionais. Bases de dados paralelas e distribuídas.

Descrição detalhada dos conteúdos programáticos

Componente Teórica

Introdução aos SGBDs: sistemas de gestão de bases de dados (SGBD); propósitos e vantagens face ao uso de ficheiros; quando não usar um SGBD; modelos interno, conceptual, e externo; independência dos dados; evolução histórica; visão geral das componentes de um SGBD; usos e mercado dos SGBDs.

Integridade em bases de dados: regras de integridade implícitas e explícitas; especificação de regras nas formas declarativa e procedimental; integridade no modelo relacional; regras de domínio, coluna, entidade, referencial, e definidas pelo utilizador.

Segurança de bases de dados: privacidade, integridade, e disponibilidade; concessão e revogação de privilégios; conjuntos de privilégios (roles); definição, processamento, e atualização de vistas (views) sobre os dados; segurança com combinação de privilégios e views.

Discos e ficheiros: tipos de memória e importância do disco rígido; níveis RAID para armazenamento distribuído; gestão de espaço em disco; gestão do tampão em memória (buffer pool); políticas de substituição de quadros; caso da inundação sequencial; registos de comprimento fixo e variável; páginas com registos de comprimento fixo e variável; ficheiros não ordenados com listas de páginas e com diretório; motivação para uso de índices; introdução aos índices por árvore B+ e por função de dispersão.

Índices: tipos de entradas de dados; índices primários e secundários; índices agrupados e desagrupados; indexação de múltiplos atributos; árvores ISAM e

B+; dispersão estática e dinâmica.

Comparação de organizações de ficheiros: modelo de custos; premissas; custo com ficheiro não ordenado e ordenado, árvore B+ agrupada, índice desagrupado de árvore B+ e de dispersão; tabela síntese.

Revisão de álgebra relacional: operadores de projecção, seleção, produto, diferença, união, interseção, junção, e outros; conversão de interrogações SQL em expressões de álgebra relacional.

Avaliação de interrogações: passos da avaliação; árvores de álgebra relacional e planos de execução; técnicas de realização de operadores relacionais; escolha do algoritmo e estatísticas; caminhos de acesso às tabelas; seletividade e fator de redução; estimação da cardinalidade e do fator de redução; geração de planos de execução; estimação do custo.

Ordenação externa: algoritmo two-way merge sort e caso geral; estimação do número de passagens e do custo da ordenação; ordenação com índices de árvore B+.

Realização de operadores relacionais: projecção com ordenação e dispersão; seleção simples com e sem índice; seleção com condições genéricas com um ou mais índices; junção com técnicas de iteração, indexação, e partição; união e agregações.

Otimização de interrogações: otimizador do Sistema R; blocos de interrogação; geração de planos alternativos; equivalências de operadores relacionais; planos com uma relação; planos com várias relações; enumeração de planos left-deep; análise do custo de planos alternativos; outras abordagens de otimização.

Desenho físico e afinação de bases de dados: avaliação da carga do sistema; importância e frequência de interrogações e atualizações; escolha de índices; uso de índices agrupados; planos só de índices; evolução do esquema conceptual e views; desnormalização; decomposição horizontal; afinação de interrogações.

Transações: propriedades de atomicidade, coerência, isolamento, e durabilidade; design de atomicidade; componentes de um gestor transaccional; utilizadores de sistemas transaccionais; flat transactions e extensões com savepoints, nested transactions, e chained transactions.

Isolamento de transações: modelos e tipos de dependência entre transações; tipos de anomalias e representação no grafo de dependências; uso de locks para evitar anomalias; transações bem formadas e bi-faseadas; registos fantasma; locks baseados em predicados e granulares; níveis de isolamento; prevenção e deteção de deadlocks; algoritmos wait-die e wound-wait; outras técnicas para isolamento, sem recurso a locks.

Recuperação de transações: abordagens steal e force e implicações no design do gestor de recuperação; algoritmo ARIES; fases de recuperação; registos no log, na tabela de transações, na tabela de páginas modificadas, e nas páginas de dados em disco; registos de checkpoint; registos no log perante operações em funcionamento normal do SGBD; recuperação em caso de crash; cenários de exemplo.

Bases de dados paralelas: motivação; terminologia do paralelismo; arquiteturas computacionais de memória partilhada, disco partilhado, e nada partilhado; paralelismo inter-interrogação, inter-operador, e intra-operador; partições de dados; paralelismo em pesquisas, ordenação, e junção de dados; otimização de interrogações.

Bases de dados distribuídas: equilíbrio entre autonomia local e transparência no acesso aos dados; aplicações locais e distribuídas; classificação segundo a heterogeneidade, autonomia, e distribuição de dados; arquiteturas cliente/servidor, servidores colaborativos, e middleware; fragmentação de dados; técnicas síncronas e assíncronas de replicação de dados; informação adicional no catálogo de sistema; interrogações distribuídas com replicação e fragmentação horizontal ou vertical de dados; junções distribuídas com semijoin e bloomjoin; controlo de concorrência com gestão de locks centralizada, com cópia primária, e distribuída; deteção de deadlocks distribuído com gestão centralizada, hierárquica, e por timeout; recuperação distribuída com o algoritmo two-phase commit; regras, vantagens, e desvantagens das bases de dados distribuídas.

Componente Teórica-Prática

Apresentação das aulas teórico-práticas: revisão de conceitos de regras declarativas; tópicos sobre desenho físico de bases de dados e avaliação de desempenho; exemplo de acesso JDBC a uma base de dados.

Linguagem PL/SQL: tipos de dados e declaração de variáveis; execução condicional e ciclos; tratamento de erros baseado em exceções; desenvolvimento de código modular; integração com a linguagem SQL; regras de integridade em forma de triggers.

Ambiente de trabalho SQL*Plus: interface PL/SQL para controlar o acesso aos dados e para garantir regras de integridade complexas; transporte de cursores e uso de funções em comandos SQL; auditoria de atividade com triggers.

Índices: pesquisa, inserção, e remoção de valores em índices de árvore B+, dispersão estática, e dinâmica; vantagens e desvantagens de cada tipo de índice; exercícios.

Acesso a bases de dados com JDBC: interrogações SQL; chamadas a procedimentos remotos; transporte e uso de cursores; acesso concorrente à base de dados.

Medição de desempenho: tempo de execução de comandos SQL com blocos de código PL/SQL; planos de execução de interrogações; comparação de custos com e sem índices auxiliares e com diferentes níveis de ocupação dos blocos em disco; estatísticas sobre tabelas e índices.

Bibliografia

Recomendada

Ramakrishnan, R., & Gehrke, J. (2003). Database management systems (3rd ed.). Boston: McGraw-Hill.

Feuerstein, S. (2009). Oracle PL/SQL programming (5th ed.). Sebastopol, CA: O'Reilly & Associates.

Gray, J., & Reuter, A. (1993). Transaction processing: Concepts and techniques. San Mateo, Calif.: Morgan Kaufmann.

Outros elementos de estudo

Coletânea com 240 exercícios de testes e exames

Gurry, M. & Corrigan, P. (1996) Oracle Performance Tuning (2nd ed.) O'Reilly.

Date, C.J. (2003) An Introduction to Database Systems (8th ed.) Addison-Wesley.

Navathe, S.B. & Elmasri, R.E. (1999) Fundamentals of Database Systems (3rd ed.). Addison Wesley

Melton, J. & Simon, A.R. (1992) Understanding the New SQL: A Complete Guide. Morgan Kaufmann

Manuais online da Oracle (SQL Language Reference, PL/SQL Language Reference, Database Error Messages, e outros)

Manuais online sobre Java (API Specification, Getting Started with JDBC, e outros)

Métodos de Avaliação

Existem dois métodos de avaliação: a) projeto (35%), 2 testes durante as aulas (27,5% + 27,5%), e participação nas aulas (10%); ou b) projeto (35%), exame (55%), e participação (10%). O projeto é avaliado em quatro etapas distintas durante o semestre. Os alunos com estatuto de trabalhador-estudante podem substituir a participação nas aulas por respostas a questões semanais.

Língua de ensino

Português