

PROGRAMA ANALÍTICO

ESTRUCTURAS DE ALMACENAMIENTO DE DATOS

Ciclo Lectivo 2006

PARTE I. Modelado conceptual de datos

1. Conceptos Generales e Introducción a los Modelos de Datos

- Terminología básica: datos, información, estructura de datos, almacenamiento de información, tipos de evocación, sistemas de base de datos.
- Objetivos del diseño de estructuras de datos. Niveles de diseño.
- Definición de modelo conceptual de datos. Características de un modelo conceptual. Modelo, esquema y ejemplar. Rol del Modelo de Datos en el ciclo vida de un sistema de información.
- Mecanismos de abstracción en Modelos de datos Conceptuales
 - Clasificación
 - Agregación
 - Generalización
 - Asociación

2. Modelo de Entidades y Relaciones Básico (MER) y Extendido (MERE)

- Componentes del Modelo de Entidades y Relaciones
 - Entidades. Instancias o ejemplares.
 - Atributos. Dominios y valores. Identificadores.
 - Relaciones. Grado y cardinalidad de una relación. Relaciones unarias y binarias.
- Componentes del Modelo de Entidades y Relaciones Extendido
 - Entidades débiles y fuertes: dependencia de identificación y dependencia de existencia.
 - Entidades asociativas.
 - Relaciones ternarias y de mayor orden.
 - Jerarquías de generalización y especialización.
 - Agregación: Relaciones entidad-relación y relación-relación.

3. Modelo relacional

- Definición y características principales.
- Elementos del modelo Relacional
 - Relaciones. Tuplas. Atributos. Dominios.
 - Esquema y Extensión.
 - Claves y superclaves. Claves candidatas, primarias y extranjeras.
 - Restricciones de integridad inherentes: de claves primarias y extranjeras.

- Concepto y Uso de Dependencias
 - Dependencias Funcionales
 - Dependencia funcional basada en claves, dependencia funcional plena y dependencia funcional transitiva. Grafo de dependencias.
 - Sistema de axiomas de Armstrong.
 - Algoritmos relativos a la normalización: Calculo de la clausura transitiva, del recubrimiento minimal, de la determinación de claves y de la verificación de descomposición sin perdida (Algoritmo de Ullman).
 - Dependencias de Inclusión: Integridad Referencial
 - Inclusión pura y basada en claves.
 - Sistema de axiomas de Casanova.
 - Acciones referenciales: reparación y rechazo.
- Operaciones básicas del álgebra relacional para manipulación de relaciones.
- Diseño utilizando mecanismos de derivación MER → esquema relacional.

4. Teoría de Normalización

- Teoría de normalización. Motivación: Problemas de actualización y de mantenimiento de integridad que provoca la redundancia.
- Descomposición con y sin pérdida. Aplicación de los algoritmos en la normalización.
- Formas normales basadas en dependencias funcionales. Primera (1FN), segunda (2FN), tercera (3FN) y de Boyce-Codd (FNBC). Algoritmos de descomposición de una relación a su equivalente en proyecciones FNBC.
- Teoría de la normalización y diseño relacional basado en Análisis.
- Introducción a Formas Normales Avanzadas: Dependencias multivalor y 4FN. Dependencias de reunión y 5FN.

PARTE II. Almacenamiento físico de Datos

1. Introducción a la organización de archivos e índices

- Almacenamiento físico de la información. Dispositivos de Almacenamiento secundario.
- Archivos. Registros. Dominio estructurante, clave de búsqueda y clave de la relación. Punteros físicos y lógicos. Operaciones sobre archivos. Organización y manejo de archivos en disco. Ordenamiento externo.
- Organización de archivos
 - Desordenado.
 - Ordenado.
 - Disperso.
- Otras organizaciones primarias de archivos: archivos en árboles B y archivos de registros mixtos.
- Índices

- Índices un solo nivel
 - Ralos y Densos.
 - Primarios.
 - De Agrupamiento.
 - Secundarios.
- Índices multinivel: estáticos y dinámicos basados en árboles B y B*.
- Índices sobre múltiples claves.
- Índices sobre estructuras de dispersión.

2. Técnicas de Dispersión

- Evocación asociativa. Características generales de las técnicas de dispersión y su uso de en memoria principal y secundaria.
- Dispersión perfecta
 - Definición. Métricas.
 - Condiciones que definen un problema apto para ser resuelto con dispersión perfecta y criterios para solucionar el incumplimiento de algunas de ellas.
- Dispersión pura
 - Conceptos principales: balde, ranura, factor de carga, sinónimos, colisión, overflow o rebalse, resolución de colisiones, tipos de clustering
 - Técnicas cerradas o estáticas
 - Open addressing ó reasignación lineal. Clustering Primario.
 - Reasignación cuadrática. Reasignación pseudoazar. Clustering secundario.
 - Reasignación realeatorizada.
 - Problemas del uso de técnicas cerradas de dispersión en memoria externa.
 - Técnicas abiertas o dinámicas
 - Hashing extensible.
 - Rebalse separado convencional.
 - Rebalse separado con crecimiento lineal.

3. Árboles n-arios

- Motivación: uso de árboles como índices multinivel dinámico y en organización primaria. Desventajas de árboles binarios de búsqueda balanceados para memoria externa. Lista de varios niveles como precursor de árboles n-arios.
- Árboles n-arios
 - Definición general. Propiedades.
 - Árboles-B.
 - Derivados de árboles B
 - Árboles B+.
 - Árboles B*.
 - Operaciones de inserción y supresión estándar y perezosas. Mecanismo de reestructuración.
- Organización ISAM.

4. Búsqueda por múltiples claves

- Evocaciones por múltiples claves y necesidad de estructuras de acceso específicas para resolverlas.
- Estructuras de datos de acceso multiclave.
 - Hashing particionado.
 - Grid-Files
 - Archivos enhebrados y ortogonales. Multilistas.
 - Archivos invertidos.
 - Bitmaps

PARTE III: Introducción a las Bases de Datos

- Características de los Sistemas de Archivos Convencionales como soporte de un Sistema de Información. Evolución de los sistemas de información basados en archivos convencionales a los sistemas de bases de datos. Ventaja del uso de base de datos.
- Características del enfoque de Base de Datos
 - Autodescripción.
 - Independencia de los datos.
 - Soporte de múltiples vistas.
 - Arquitectura de 3 niveles
 - Ciclo de vida.