

Структуры баз данных

Lecture Notes N 9 v.[2+]

Б. Мишнев



Понятие файла

- Совокупность связанных записей (логических или физических), рассматриваемых как единое целое
- Совокупность данных, находящихся во внешней памяти и доступных программе

Файловые системы

- Первая развитая файловая система была разработана фирмой IBM для ее серии 360.
- В этой системе поддерживались как чисто последовательные, так и индексно-последовательные файлы

Последовательные файлы

- Доступ к файлу осуществляется последовательно от начала к концу
- Записи обрабатываются в порядке их считывания
- Необходимо контролировать достижение конца этого файла (EOF)

Текстовые файлы

- Последовательный файл, в котором каждая логическая запись состоит из одного «печатаемого символа»
- «Печатаемый символ» - двоичная битовая комбинация, представляющая клавишу традиционной клавиатуры (включая символы «возврат каретки» и «перевод строки»)
- Кодировки ASCII и Unicode

Индексный файл

- Состоит из списка элементов, каждый из которых содержит значение поля ключа, за которым следует указание о месторасположении данной записи
- Индекс обеспечивает прямой доступ к отдельным записям файла по значению ключа

Хеширование

Метод, обеспечивающий прямой доступ к записям без использования каких-либо других структур

- Пространство хранения файла делится на сектора (buckets)
- Записи распределяются по секторам согласно алгоритму хеширования, преобразующему значения ключа в номер сегмента

Область применения файлов

Файловые системы обычно обеспечивают хранение слабо структурированной информации, оставляя дальнейшую структуризацию прикладным программам.

Викторина 1

- Нарисуйте в конспекте таблицу

№ вопр.	Ответ	Проверка
1.1		
1.2		
1.3		
1.4		
1.5		

- В качестве ответов используйте:
I – да, - нет, O – не известно

Вопрос 1.1

- Правильно ли то, что процедура создания дескриптора файла называется открытием файла?

Вопрос 1.2

- Правильно ли то, что EOF – это аббревиатура от слов «таблица размещения файла»?

Вопрос 1.3

- Верно ли, что FAT обозначает метку «конец файла»?

Вопрос 1.4

- Верно ли, хэширование обеспечивает прямой доступ к записям без использования каких-либо других структур?

Вопрос 1.5

- Верно ли, что в большинстве языков программирования периферийные устройства компьютера (например, клавиатура) рассматриваются как последовательные файлы?

Викторина 1 (ответы)

- Сравните свои ответы и отметьте правильные в графе проверка «галочкой»

№ вопроса	Ответ	Проверка
1.1	I	
1.2	-	
1.3	-	
1.4	I	
1.5	I	

- Количество правильных ответов запишите!

База данных

- Набор данных, многомерный в смысле наличия внутренних связей между его отдельными элементами, что позволяет получать доступ к информации, исходя из различных точек зрения
- Совокупность данных, обеспечивающая адекватную модель предметной области при минимальной избыточности информации

Основные функции СУБД

- Управление данными во внешней памяти;
- Управление буферами оперативной памяти;
- Управление транзакциями;
- Журнализация и восстановление БД после сбоев;
- Поддержание языков БД.

БД, основанные на инвертированных списках

- a. Строки таблиц упорядочены системой в некоторой физической последовательности.
- b. Физическая упорядоченность строк всех таблиц может определяться и для всей БД (так делается, например, в Datacom/DB).
- c. Для каждой таблицы можно определить произвольное число ключей поиска, для которых строятся индексы. Эти индексы автоматически поддерживаются системой, но явно видны пользователям
- d. Общие правила определения целостности БД отсутствуют.

Иерархические БД (IMS)

- Состоят из упорядоченного набора деревьев
- Тип дерева в целом представляет собой иерархически организованный набор типов записи.
- Все экземпляры данного типа потомка с общим экземпляром типа предка называются близнецами.
- Для БД определен полный порядок обхода - сверху-вниз, слева-направо
- Автоматически поддерживается целостность ссылок между предками и потомками

Сетевые БД (IDMS)

- Сетевой подход является расширением иерархического - потомок может иметь любое число предков.
- Состоит из набора записей и набора связей между этими записями
- Ограничение целостности по ссылкам

Достоинства реляционного подхода

- Наличие небольшого набора абстракций;
- Наличие простого и в то же время мощного математического аппарата;
- Возможность ненавигационного манипулирования данными без необходимости знания конкретной физической организации баз данных во внешней памяти.

Объектно-ориентированные базы данных

Базируется на следующих концепциях:

- Объекта и идентификатора объекта;
 - Атрибутов и методов;
 - Классов;
 - Иерархии и наследования классов.
- Примеры: O2, ORION, GemStone и Iris.

Распределенные базы данных

- База данных, которая разделена между несколькими компьютерами в сети и которую можно совместно использовать несколькими пользователями
- База данных физически распределяется по узлам данных при помощи **фрагментации** и **репликации** («прозрачно» для пользователя)
- Примеры: DNS, Oracle, Lotus Notes

Викторина 2

- Нарисуйте в конспекте таблицу

№ вопр.	Ответ	Проверка
2.1		
2.2		
2.3		
2.4		
2.5		

- В качестве ответов используйте:
I – да, - нет, O – не известно

Вопрос 2.1

- Правильно ли то, что в Латвийской республике использование баз данных, содержащих персональную информацию регламентирует “Fizisko personu datu aizsardzības likums”?

Вопрос 2.2

- Правильно ли то, что распределенные базы данных могут содержать как фрагментированные, так и реплицированные данные?

Вопрос 2.3

- Верно ли, что в реляционных базах данных информация отображается в виде прямоугольных таблиц, называемых «отношениями» (relation)?

Вопрос 2.4

- Верно ли, что строка в таблице реляционной базы данных называется «кортежем» (tuple)?

Вопрос 2.5

- Верно ли, что столбцы в отношении именуются «атрибутами» (attribute)?

Викторина 2 (ответы)

- Сравните свои ответы и отметьте правильные в графе проверка «галочкой»

№ вопроса	Ответ	Проверка
2.1	I	
2.2	I	
2.3	I	
2.4	I	
2.5	I	

- Количество правильных ответов запишите!

Проектирование реляционных БД

- Каким образом отобразить объекты предметной области в абстрактные объекты модели данных - **логическое проектирование.**
- Как обеспечить эффективность выполнения запросов к базе данных - **физическое проектирование.**

Процесс проектирования

- Принятие решений о том
 - из каких отношений должна состоять БД и
 - какие атрибуты должны быть у этих отношений.
- Нормализации схем отношений, причем каждая следующая нормальная форма обладает свойствами лучшими, чем предыдущая (1-3 NF, BCNF, 5NF-PJ/NF).

Транзакции

- Под транзакцией понимается неделимая с точки зрения воздействия на БД последовательность операторов манипулирования данными (чтения, удаления, вставки, модификации) такая, что либо результаты всех операторов, входящих в транзакцию, отображаются в БД («**все**»), либо воздействие всех этих операторов полностью отсутствует («**ничего**»).

Журнализация изменений БД

- *Цель: обеспечение возможности восстановления согласованного состояния базы данных после любого сбоя.*
- **Общими принципами восстановления являются следующие:**
 - результаты зафиксированных транзакций должны быть сохранены в восстановленном состоянии базы данных;
 - результаты незафиксированных транзакций должны отсутствовать в восстановленном состоянии базы данных.

Язык реляционных баз данных SQL

- Разработан IBM в рамках проекта экспериментальной реляционной СУБД System R в середине 70-х (SEQUEL -Structured English Query Language)
- В настоящее время SQL реализован практически во всех коммерческих реляционных СУБД (DB2, Oracle, Informix etc.) – стандарт с 1989 г.

Возможности SQL

- Формулирование запросов и манипулирование БД
- Определение схемы БД
- Определение ограничений целостности БД
- Возможности определения структур физического уровня
- Авторизация доступа к отношениям и их полям, точкам сохранения транзакции и откатов и др.

Викторина 3

- Нарисуйте в конспекте таблицу

№ вопр.	Ответ	Проверка
3.1		
3.2		
3.3		
3.4		
3.5		

- В качестве ответов используйте:
I – да, - нет, O – не известно

Вопрос 3.1

- Правильно ли то, что проектирование реляционных баз данных сводится к разработке отношений, входящих в эту базу данных?

Вопрос 3.2

- Правильно ли то, что приведение набора составляющих базу данных отношений к очередной нормальной форме позволяет сделать базу данных более удобной в использовании?

Вопрос 3.3

- Верно ли, что операция JOIN предназначена для объединения двух отношений?

Вопрос 3.4

- Верно ли, что в результате выполнения операции `SELECT` будет создано новое отношение (новая таблица) состоящая из отобранных столбцов (атрибутов) исходного отношения?

Вопрос 3.5

- Верно ли, что язык SQL (Structured Query Language – язык структурных запросов) был разработан фирмой IBM?

Викторина 3 (ответы)

- Сравните свои ответы и отметьте правильные в графе проверка «галочкой»

№ вопроса	Ответ	Проверка
3.1	I	
3.2	I	
3.3	I	
3.4	-	
3.5	I	

- Количество правильных ответов запишите!

Производственные информационные системы

- Системы обработки транзакций (transaction processing systems - TPS).
- Системы управления процессом (process control systems - PCS).
- Системы автоматизации делопроизводства (office automation systems - OAS).
- Системы генерации отчетов (information reporting systems - IRS)

Системы аналитической обработки данных в режиме реального времени (OLAP)

- OLAP-системы обеспечивают решение аналитических задач: анализ ключевых показателей деятельности, маркетинговый и финансово-экономический анализ, анализ сценариев, моделирование, прогнозирование и т.д. Такие системы не обусловлены особенностями информационной инфраструктуры компании и могут работать со всеми необходимыми данными, независимо от их источников.

Корпоративное хранилище данных (Data Warehouse)

- Достоинством предлагаемого решения является возможность хранить практически **неограниченный объем данных**.
- Хранилище данных **позволяет расширить возможности использования информации** за счёт ускоренного поиска, разграничения прав доступа сотрудников к данным, управления жизненным циклом информации.
- Программные средства, включённые в состав решения, **поддерживают большинство форматов данных** (фото, видео, аудиозаписи, электронные документы в форматах т.д.)

Системы поддержки принятия решений (СППР, DSS)

- Интерактивные компьютерные информационные системы, которые используют модели решений и специализированные базы данных для помощи менеджерам в принятии управленческих решений.

Литература

- Дж.Г.Брукшир. Введение в компьютерные науки. «Вильямс», 2001, с. 433 – 504.
- Введение в системы баз данных
Дейт К. Дж., 2001 г., Изд.: Вильямс, 1072
стр.

Спасибо за внимание!

Dr. Sc Ing. Борис Мишнев

