

Введение в КОМПЬЮТЕРНЫЕ НАУКИ

ЛЕКТОР К.Т.Н. МОХОВ В.А.

ГЛАВА 9. СТРУКТУРЫ БАЗ ДАННЫХ

Часть 9: Структуры баз данных

- ▶ 9.1 Общие понятия
- ▶ 9.2 Реляционная модель
- ▶ 9.3 Объектно-ориентированные базы данных
- ▶ 9.4 Обеспечение целостности баз данных
- ▶ 9.5 Традиционные структуры файлов
- ▶ 9.6 Сбор данных
- ▶ 9.7 Влияние технологий баз данных на общество

База данных

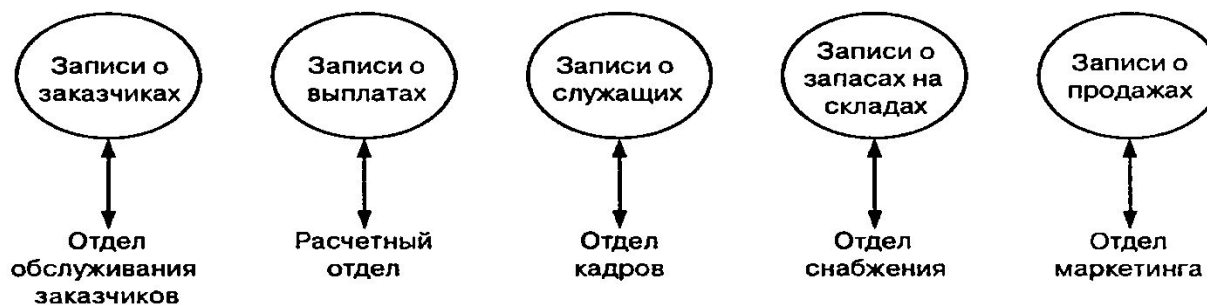
Многомерный набор данных (в смысле наличия внутренних связей между отдельными его элементами), позволяющий получить доступ к информации, исходя из различных точек зрения.

Рисунок 9.1 Структура баз данных в сравнении с

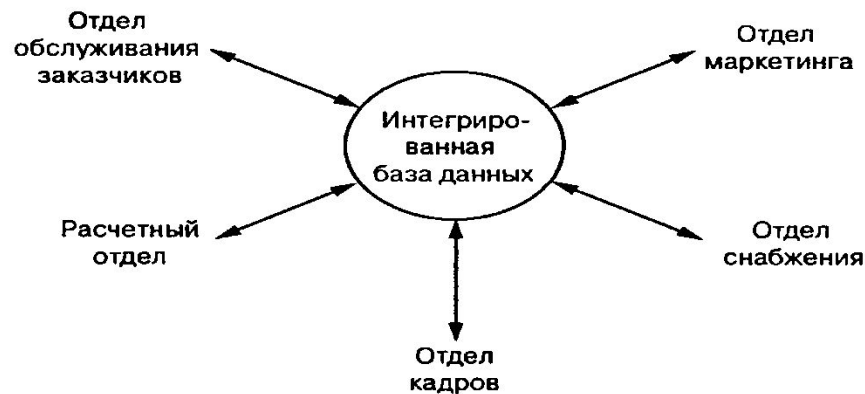
9-3

файловой структурой

а) Информационная система, ориентированная на использование плоских файлов



б) Информационная система, ориентированная на использование базы данных



Построение обычной файловой системы и системы базы данных

Рисунок 9.2 Концептуальные уровни системы базы данных

9-4



Концептуальные уровни системы базы данных

Схема

- ▶ **Схема** : представляет собой полное описание структуры базы данных, которое используется ее программным обеспечением для обслуживания базы данных в целом.
- ▶ **Подсхема** : это описание части базы данных, соответствующей нуждам отдельного пользователя.

Системы управления базой данных

- ▶ Система Управления базой данных (СУБД): слой программного обеспечения, который управляет базой данных в ответ на запросы из приложения
- ▶ **Распределенная база данных** : База данных хранится на нескольких машинах

СУБД будет маскировать эту организационную деталь от своих пользователей

- ▶ **Независимость данных** : Возможность изменить организацию базы данных без изменения прикладного программного обеспечения, использующего его

Модели баз данных

- ▶ **Модель базы данных:** концептуальное представление базы данных
 - ▶ Реляционная модель
 - ▶ Объектно-ориентированные базы данных

Реляционная модель базы данных

6-0

- ▶ **Отношения:** прямоугольная таблица
 - ▶ **Атрибуты:** Столбец в таблице
 - ▶ **Кортеж:** Строка в таблице

Рисунок 9.3 Отношение (таблица, relation), содержащее сведения о сотрудниках

EmplId	Name	Address	SSNum
25X15	Джо Бейкер	ул. Верхняя, 33	111223333
34Y70	Шери Кларк	ул. Верхняя, 33	999009999
23Y34	Джерри Смит	пер. Круглый, 15	111005555
.	.	.	.
.	.	.	.
.	.	.	.

Отношение, содержащее сведения о сотрудниках

Реляционное проектирование

- ▶ Избегайте нескольких концепций в одном отношении
- ▶ Может привести к лишним данным
 - ▶ Удаление кортежа могло также удалить необходимую но несвязанную информацию

Улучшение реляционного дизайн

- ▶ **Разложение:** Разделение столбцов отношения в двух или более отношений, дублируя эти столбцы, необходимые для поддержания отношений
- ▶ **Декомпозиция без потерь:** «Корректная декомпозиция, при котором не теряется какая-либо информация»

Рисунок 9.4 Отношение, содержащее избыточную информацию

EmpId	Name	Address	SSN	JobId	JobTitle	SkillCode	Dept	StartDate	TermDate
25X15	Джо Бейкер	ул. Верхняя, 33	111223333	F5	Начальник группы	FM3	Отдел сбыта	9-1-1998	9-30-1999
25X15	Джо Бейкер	ул. Верхняя, 33	111223333	D7	Начальник отдела	D2	Отдел сбыта	10-1-1999	*
34Y70	Шери Кларк	ул. Верхняя, 33	999009999	F5	Начальник группы	FM3	Отдел сбыта	10-1-1998	*
23Y34	Джерри Смит	пер. Круглый, 15	111005555	S25X	Секретарь	T5	Отдел кадров	3-1-1996	4-30-1998
23Y34	Джерри Смит	пер. Круглый, 15	111005555	S25Z	Секретарь	T6	Бухгалтерия	5-1-1998	*
.
.
.

Отношение, содержащее избыточную информацию

Рисунок 9.5 База данных с информацией о сотрудниках, состоящая из трех отношений

Отношение EMPLOYEE			
EmplId	Name	Address	SSNum
25X15	Джо Бейкер	ул. Верхняя, 33	111223333
34Y70	Шери Кларк	ул. Верхняя, 33	999009999
23Y34	Джерри Смит	пер. Круглый, 15	111005555
.	.	.	.
.	.	.	.
.	.	.	.

Отношение JOB			
JobId	JobTitle	SkillCode	Dept
S25X	Секретарь	T5	Отдел кадров
S26Z	Секретарь	T6	Бухгалтерия
F5	Начальник группы	FM3	Отдел сбыта
.	.	.	.
.	.	.	.
.	.	.	.

Отношение ASSIGNMENT			
EmplId	JobId	StartDate	TermDate
23Y34	S25X	3-1-1996	4-30-1998
34Y70	F5	10-1-1998	*
23Y34	S25Z	5-1-1998	*
.	.	.	.
.	.	.	.
.	.	.	.

База данных с информацией о сотрудниках, состоящая из трех отношений

Рисунок 9.6 Поиск отделов в которых работал сотрудник 23Y34

Отношение EMPLOYEE

EmpId	Name	Address	SSNum
25X15	Джо Бейкер	ул. Верхняя, 33	111223333
34Y70	Шери Кларк	ул. Верхняя, 33	999009999
23Y34	Джерри Смит	пер. Круглый, 15	111005555
.	.	.	.
.	.	.	.

Отношение JOB

JobId	JobTitle	SkillCode	Dept
S25X	Секретарь	T5	Отдел кадров
S26Z	Секретарь	T6	Бухгалтерия
F5	Начальник группы	FM3	Отдел сбыта
.	.	.	.
.	.	.	.

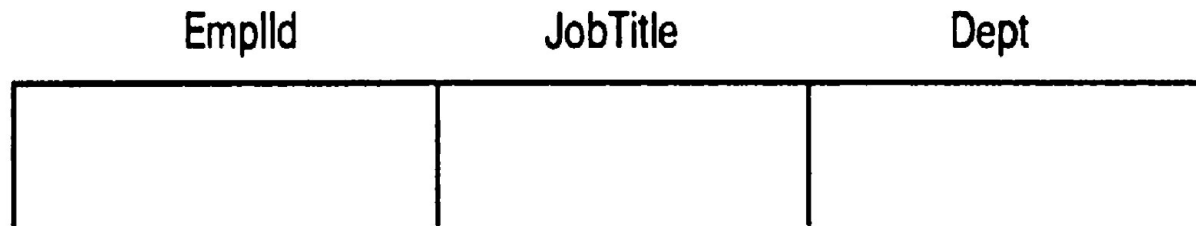
Содержатся в отделе кадров и бухгалтерии

Отношение ASSIGNMENT

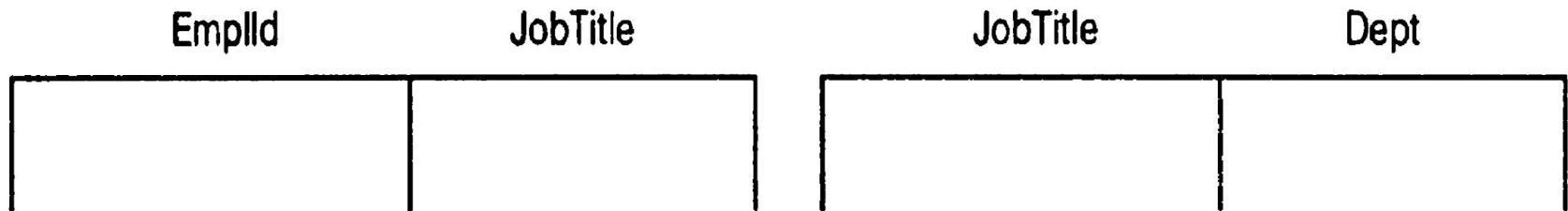
EmpId	JobId	StartDate	TermDate
23Y34	S25X	3-1-1996	4-30-1998
34Y70	F5	10-1-1998	*
23Y34	S25Z	5-1-1998	*
.	.	.	.
.	.	.	.

Задания, проведенные работником 23Y34

Рисунок 9.7 Отношение и декомпозиция без потерь



Отношение, содержащее три атрибута: EmpId, JobTitle и Dept



Два отношения, содержащие атрибуты EmpId, JobTitle и JobTitle, Dept

Реляционные операции



- ▶ **Select:** Выбор строки
- ▶ **Project:** Выбор столбца
- ▶ **Join:** Сбор информации от двух или более отношений

Рисунок 9.8 Операция Select (employee – наемный рабочий)

Отношение EMPLOYEE

EmpId	Name	Address	SSN
25X15	Джо Бейкер	ул. Верхняя, 33	111223333
34Y70	Шери Кларк	ул. Верхняя, 33	999009999
23Y34	Джерри Смит	пер. Круглый, 15	111005555
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

NEW ← SELECT from EMPLOYEE where EmpId = "34Y70"

Отношение NEW

EmpId	Name	Address	SSN
34Y70	Шери Кларк	ул. Верхняя, 33	999009999

Операция SELECT

Рисунок 9.9 Операция Project

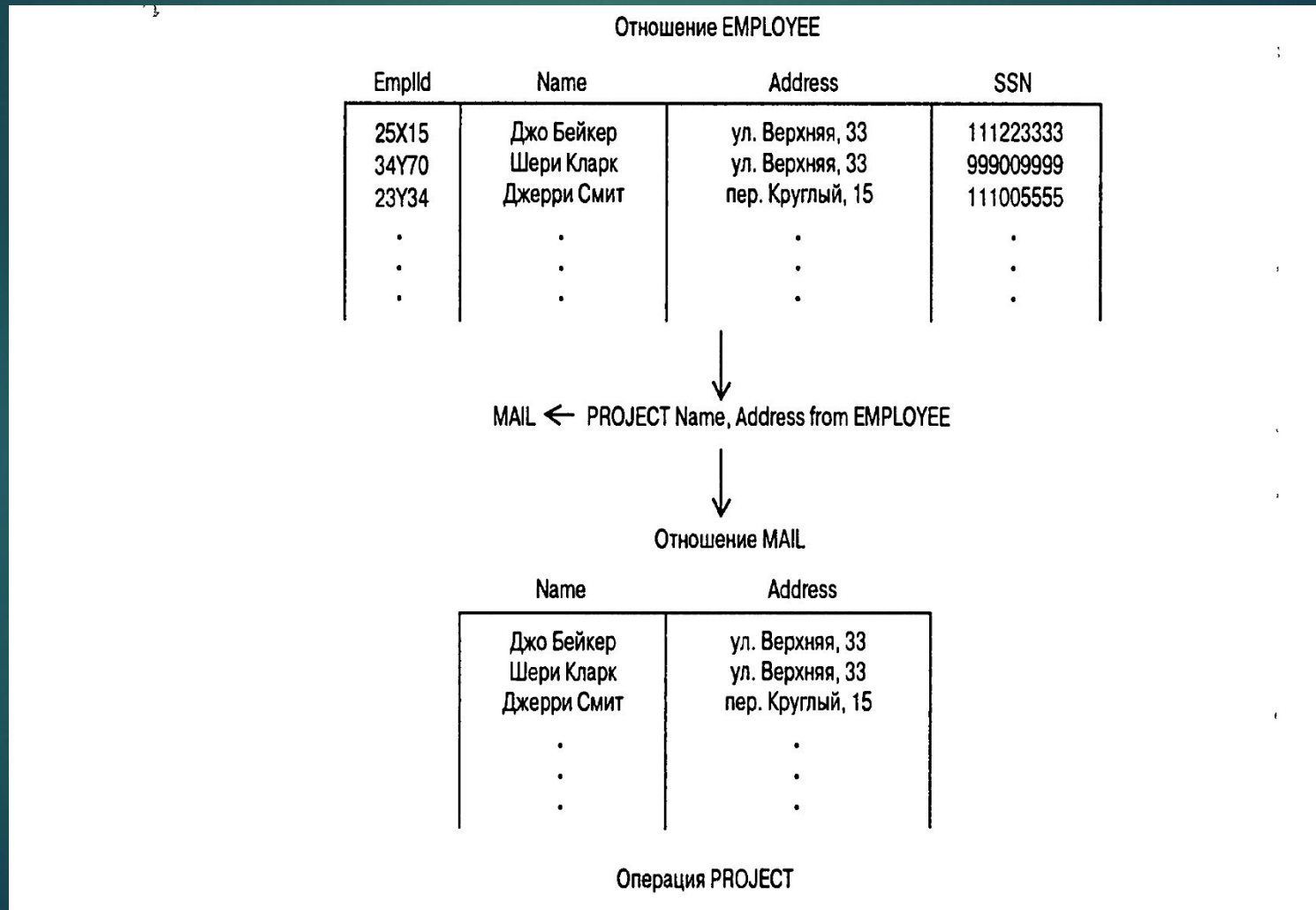


Рисунок 9.10 Операция Join

Отношение A

V	W
r	2
t	4
p	6

Отношение B

X	Y	Z
5	g	p
4	d	e
2	m	q
4	t	f

$C \leftarrow \text{JOIN A and B where } A.W = B.X$

Отношение C

A.V	A.W	B.X	B.Y	B.Z
r	2	2	m	q
t	4	4	d	e
t	4	4	t	f

Операция JOIN

Рисунок 9.11 Другой пример применения операции Join

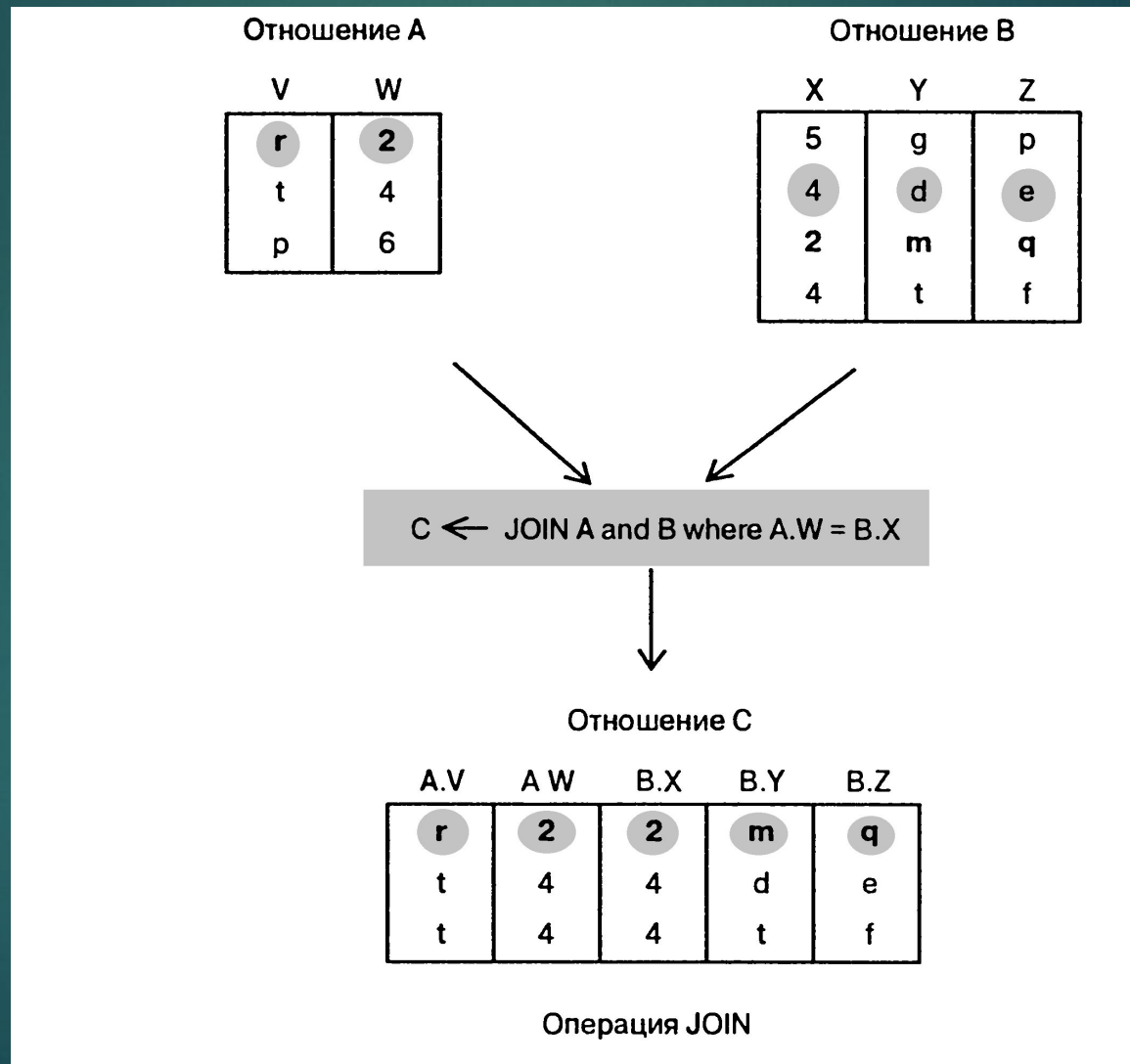


Рисунок 9.12 Применение операции Join assignment – назначение, job – работа

Отношение ASSIGNMENT				Отношение JOB			
Empl Id	Job ID	Start Date	Term Date	Job Id	Job Title	Skill Code	Dept
23Y34	S25X	3-1-1996	4-30-1998	S25X	Секретарь	T5	Отдел кадров
34Y70	F5	10-1-1998	*	S26Z	Секретарь	T6	Бухгалтерия
25X15	S26Z	5-1-1998	*	F5	Начальник группы	FM3	Отдел сбыта
.
.
.

NEW1 ← JOIN ASSIGNMENT and JOB where ASSIGNMENT JobId = JOB JobId

↓

Отношение NEW1

ASSIGNMENT EmplId	ASSIGNMENT. JobID	ASSIGNMENT. StartDate	ASSIGNMENT TermDate	JOB JobId	JOB JobTitle	JOB. SkillCode	JOB Dept
23Y34	S25X	3-1-1996	4-30-1998	S25X	Секретарь	T5	Отдел кадров
34Y70	F5	10-1-1998	*	F5	Начальник группы	FM3	Отдел сбыта
25X15	S26Z	5-1-1998	*	S26Z	Секретарь	T6	Бухгалтерия
.
.
.

Пример использования операции JOIN

Язык SQL (Structured Query Language)

- ▶ Операция для управления кортежами
 - ▶ insert
 - ▶ update
 - ▶ delete
 - ▶ select

SQL Примеры

- ▶

```
select EmpId, Dept
from ASSIGNMENT, JOB
where ASSIGNMENT.JobId =
JOB.JobId
      and ASSIGNMENT.TermData = "*"

```
- ▶

```
insert into EMPLOYEE
values ('43212', 'Sue A. Burt',
      '33 Fair St.', '444661111')

```

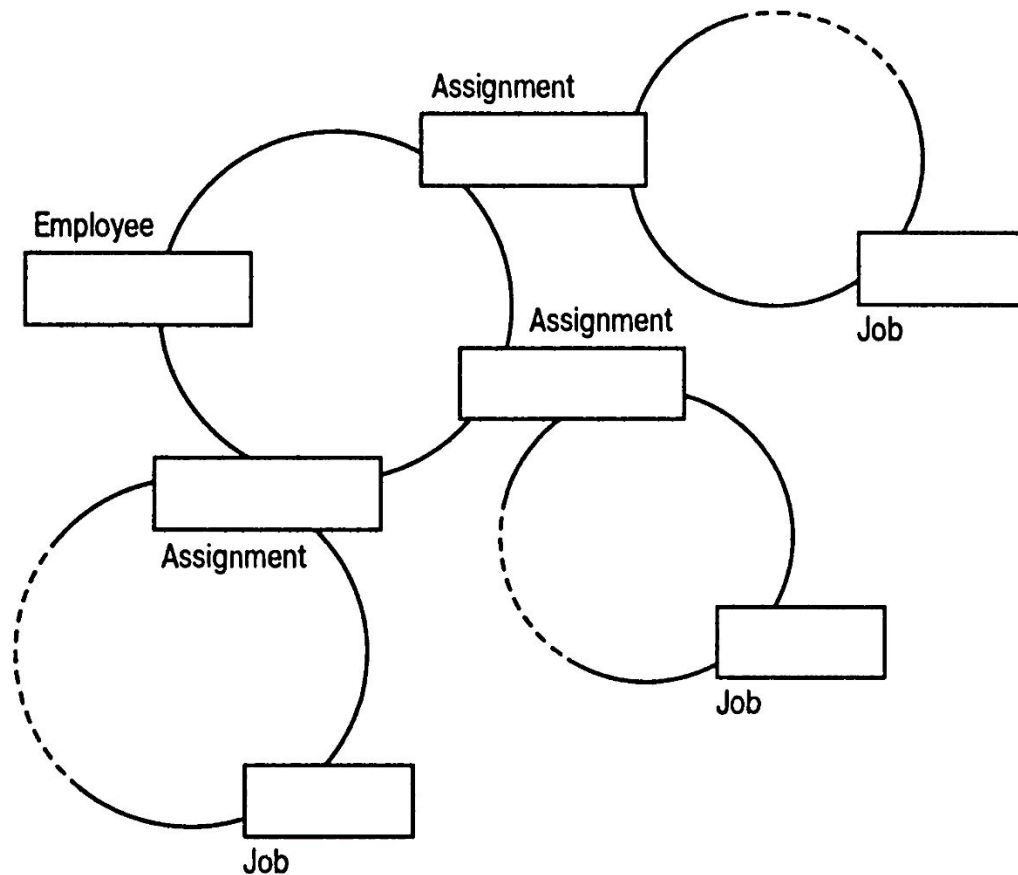

SQL Примеры (продолжение)

- ▶ `delete from EMPLOYEE
where Name = 'G. Jerry Smith'`
- ▶ `update EMPLOYEE
set Address = '1812 Napoleon
Ave.'
where Name = 'Joe E. Baker'`

Объектно-ориентированные базы данных

- ▶ **Объектно-ориентированная база данных:** База данных строится с применением объектно-ориентированной парадигмы
 - ▶ Каждый элемент хранится в виде объекта
 - ▶ Отношения указывают связи между объектами
 - ▶ СУБД поддерживает межобъектные ссылки

Рисунок 9.13 Связи между объектами в объектно-ориентированной базе данных



Связи между объектами в объектно-ориентированной базе данных

Преимущества объектно-ориентированных баз данных



- ▶ Сочетает парадигмы ООП
- ▶ В обработчик данных может быть встроен интеллект
- ▶ Может обрабатывать экзотические типы данных

Пример: multimedia

Обеспечение целостности баз данных

- ▶ **Протокол фиксации/отката изменений**
 - ▶ Пример: перевод денег между банковскими счетами
 - ▶ **Журнал транзакций:** содержит сведения о каждом действии выполняемой транзакции
 - ▶ **Фиксация точки:** точка, в которой сделка была записана в журнале
 - ▶ **Откат:** Процесс отката транзакции

Обеспечение целостности баз данных (продолжение)

- ▶ Одновременные проблемы доступа
 - ▶ Проблема недостоверных итогов
 - ▶ Проблема потерянного обновления
- ▶ **Блокировка** = предотвращение несанкционированного доступа при транзакции
 - ▶ **Разделяемая** блокировка: используется при чтении данных
 - ▶ **Эксклюзивная** блокировка: используется при изменении данных

Последовательные файлы

- ▶ **Последовательный файл:** Файл содержание которого может быть прочитано только по порядку
 - ▶ Читатель должен быть в состоянии обнаружить конец файла (EOF)
 - ▶ Данные могут храниться в логических записях, сортированных ключевым полем
 - ▶ Значительно увеличивает скорость пакетных обновлений

Рисунок 9.14 Структура простого последовательного файла реализованного как текстовый файл

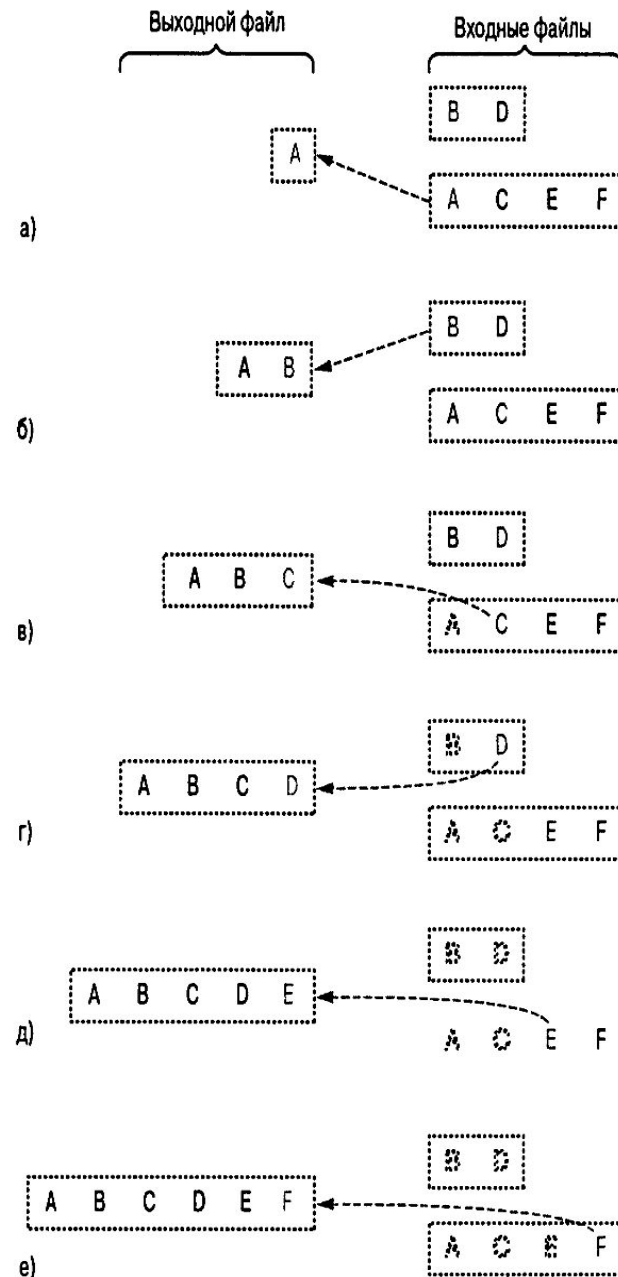


Рисунок 9.15 Процедура слияния двух последовательных файлов

```
procedure MergeFiles (InputFileA, InputFileB, OutputFile)
if (both input files at EOF) then (Stop, with OutputFile empty)
if (InputFileA not at EOF) then (Declare its first record to be its current record)
if (InputFileB not at EOF) then (Declare its first record to be its current record)
while (neither input file at EOF) do
    (Put the current record with the "smaller" key field value in OutputFile;
    if (that current record is the last record in its corresponding input file)
        then (Declare that input file to be at EOF)
        else (Declare the next record in that input file to be the file's current record)
    )
```

Starting with the current record in the input file that is not at EOF,
copy the remaining records to OutputFile.

Рисунок 9.16
Применение
алгоритма слияния
(буквы
использованы для
представления
содержимого
записей , каждая
буква
представляет
собой значение
поля ключа
соответствующей
записи.)



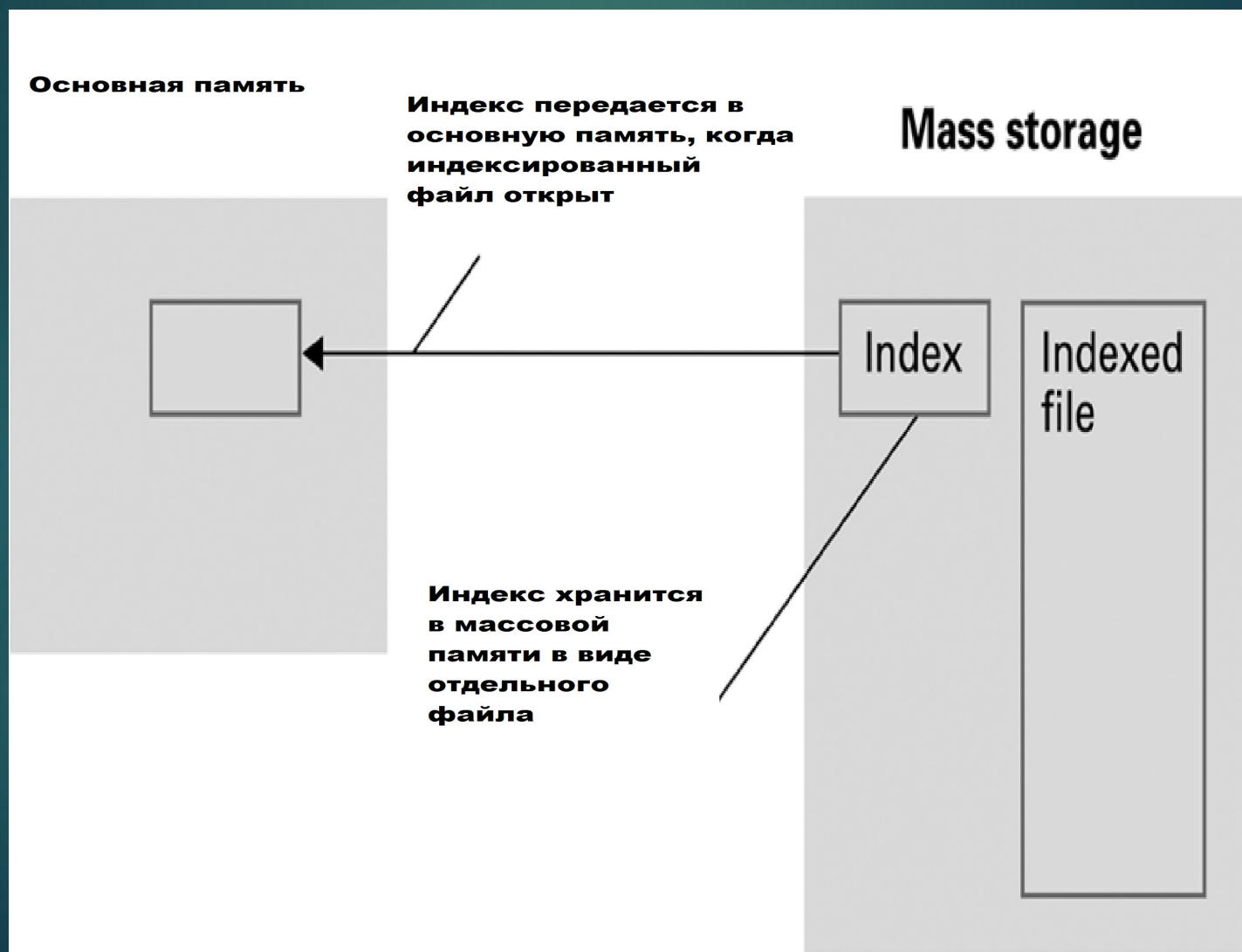
Применение алгоритма слияния (буквы использованы для представления содержимого записей, каждая буква представляет собой значение поля ключа соответствующей записи)

Индексируемые файлы



- ▶ **Index:** Список значений ключа и местоположение их связанных отчетов

Рисунок 9.17 Открытие индексированного файла



Хэширование

- ▶ Каждая запись имеет ключевое поле
- ▶ Пространство для хранения делится на сегменты
- ▶ Хэш-функция вычисляет номер сегмента для каждого значения ключа
- ▶ Каждая запись хранится в сегменте, соответствующей хэш его ключом

Рисунок 9.18 Перемещение ключевое значение поля 25X3Z к одному из 41 сегмента

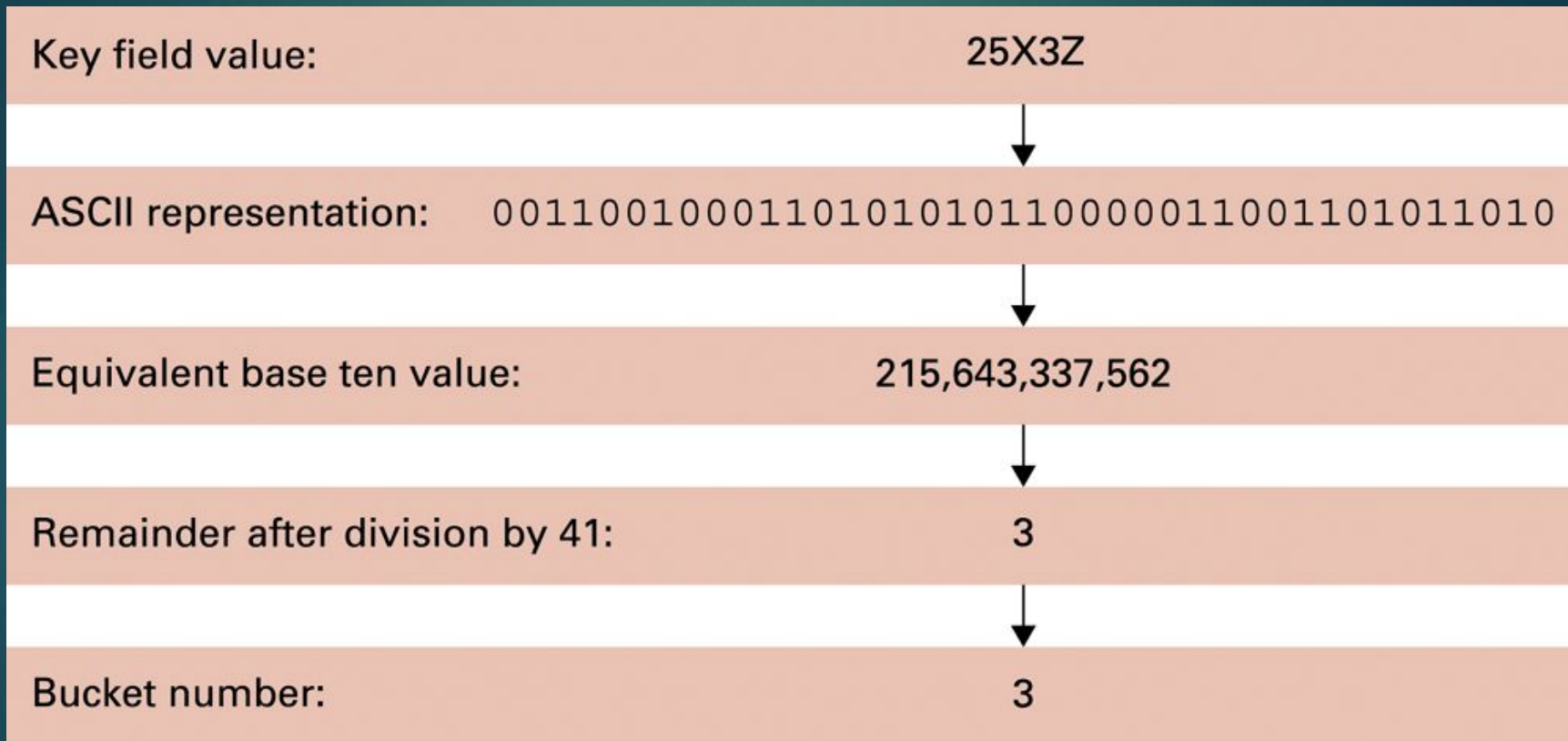
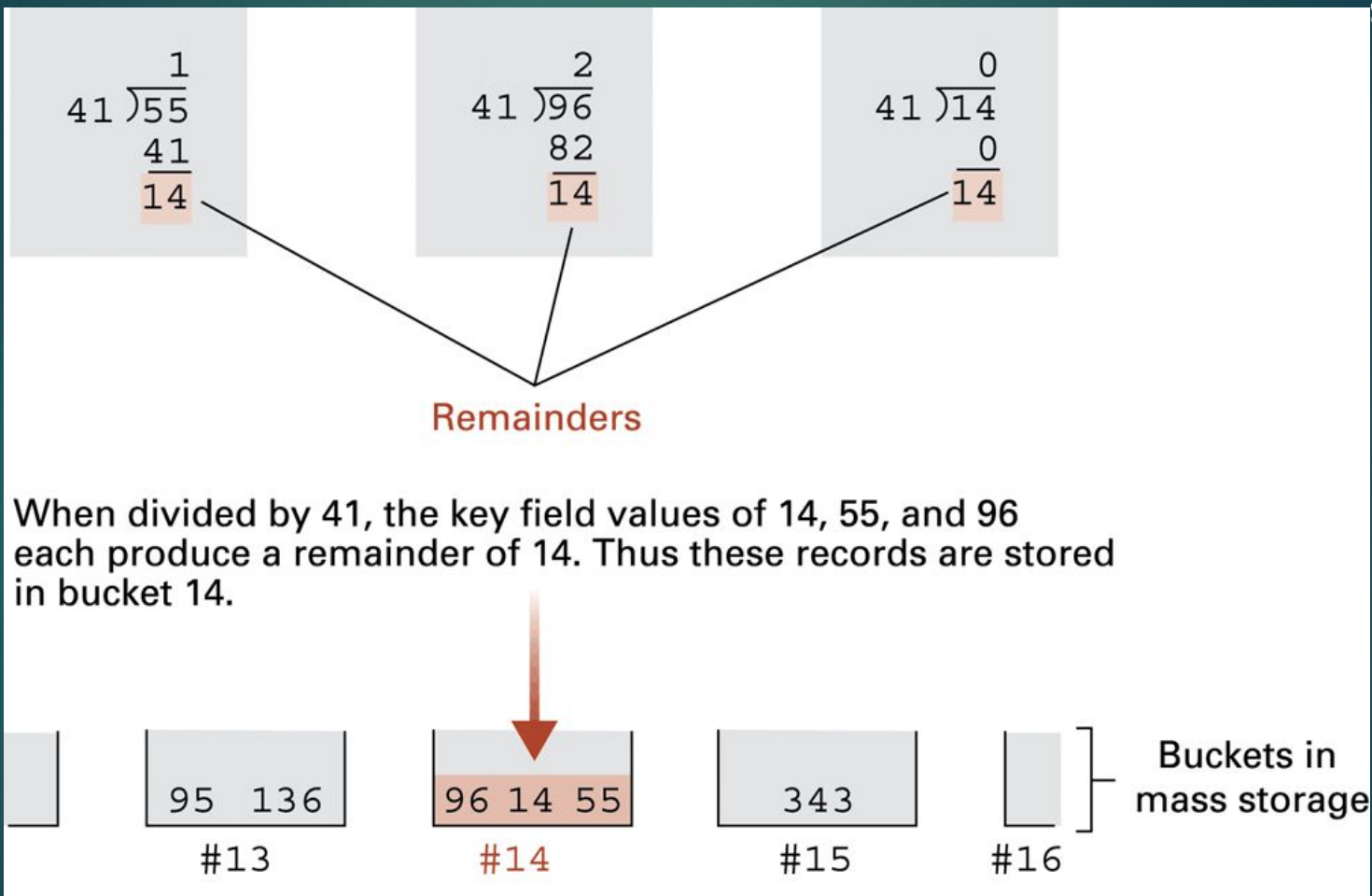


Рисунок 9.19 Рудименты системы хеширования

0-39



Коллизии в хешировании

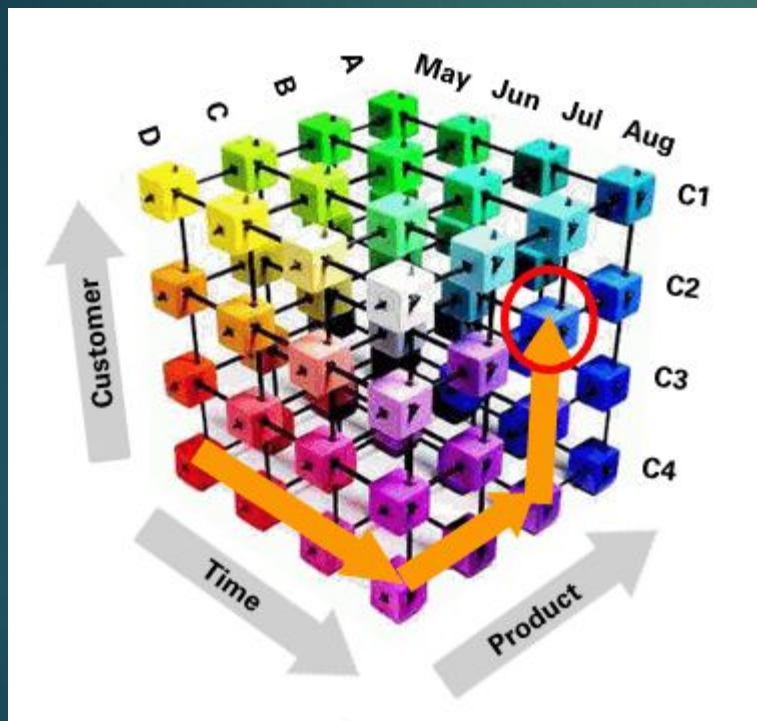
- ▶ **Коллизия:** Ситуация, когда несколько записей попадает в один и тот же сегмент
 - ▶ Главная проблема, переполнение таблицы более 75%
 - ▶ Решение: увеличить количество сегментов и переделывать все данные

Data mining

(«ДОБЫЧА ДАННЫХ»)

- ▶ **«Добыча данных»:** Область информатики, которая имеет дело с обнаружением образцов в коллекциях данных
- ▶ **Информационное хранилище :** Статический набор данных, который будет добыт
 - ▶ **Куб данных :** Данные, представленные со многих точек зрения, с целью «добычи»

Кубы данных



	Март	Февраль	Январь
	США	Канада	Мексика
Напитки	10 000	2000	1 000
Продукты питания	5000	500	250
Прочие товары	5000	500	250

Стратегия «добычи данных»

- ▶ Описание класса
- ▶ Распознавание класса
- ▶ Кластерный анализ
- ▶ Ассоциативный анализ
- ▶ Анализ изолированной части
- ▶ Последовательный анализ картины

Влияние технологий баз данных на общество



- ▶ Проблемы
 - ▶ Сбор огромного количества личных данных
 - ▶ Часто без знания или значимого согласия пострадавших
 - ▶ Объединение данных производит новый, более инвазивной информации
 - ▶ Ошибки широко распространены и трудно исправить
 - ▶ Средства
 - ▶ применении ограничений правового характера.
 - ▶ Отрицательная реклама может быть более эффективной