

Базовые понятия и технологии управления данными



В ходе развития информационных систем были сформулированы принципы организации больших массивов данных:

- принцип *интеграции данных*, в соответствии с которым все данные накапливаются и хранятся централизованно, образуя динамически обновляемую модель предметной области;
- принцип *независимости прикладных программ от данных*, т.е. отделения логической модели данных от средств управления ими.

База данных (БД) представляет собой данные, организованные и обрабатываемые в накопителях в соответствии с определенными правилами хранения и доступа.

Фиксированная, строго оговоренная структура хранения данных и их безусловная типизация отличает базу данных от текстовых и табличных процессоров, а широкая гамма допустимых операций на множествах является важным преимуществом ее перед пакетами прикладных программ и системами программирования.

Требования, предъявляемые к базам данных

Множественное использование данных

Простота

Легкость использования

Гибкость использования

Быстрая обработка запросов на данные

Язык взаимодействия конечных пользователей с системой

База данных—это основа для будущего наращивания прикладных программ

Сохранение затрат умственного труда

Наличие интерфейса прикладного программирования

Распределенная обработка данных

Адаптивность и расширяемость

Контроль за целостностью данных

Восстановление данных после сбоев

Вспомогательные средства

Автоматическая реорганизация и перемещение

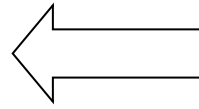
Информационная база

Коллекция
записей
данных

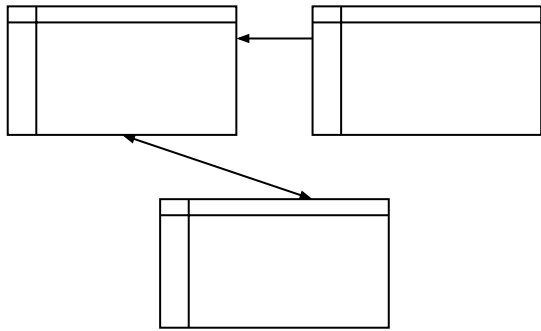
Описания
данных —
метаданные



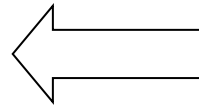
Концептуальный уровень



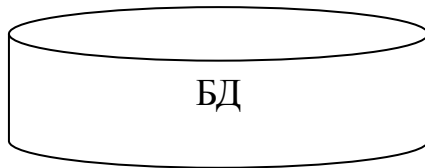
Пользователь
Разработчик ИС



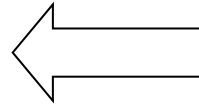
Логический (внешний) уровень



Прикладной
программист



Физический (внутренний) уровень



СУБД

Уровни представления данных

Два класса языков работы с данными

Для выражения обобщенного взгляда на данные применяют язык описания данных (ЯОД, DDL - Data Definition Language) внутреннего уровня, включаемый в состав СУБД. Описание представляет собой модель данных и их отношений, т. е. структур, из которых образуется БД.

Языки манипулирования данными (ЯМД, DML - Data Manipulating Language) обычно включает в себя средства запросов к базе данных и поддержания базы данных (добавление, удаление, обновление данных, создание и уничтожение БД, изменение определений БД, обеспечение запросов к справочнику БД).

Качественные характеристики языков запросов

*Селективная
мощность*

*Простота
изучения*

*Уровень
процедурности*

*Модульность
построения
языка*

Наиболее распространенным языком для работы с базами данных является SQL (Structured Query Language), в своих последних реализациях предоставляющий не только средства для спецификации и обработки запросов на выборку данных, но также и функции по созданию, обновлению, управлению доступом и т. д.

Пример SQL-запроса

Покажи мне всех студентов, которые получили оценку «отлично» по дисциплине «математика»

```
SELECT * FROM STUDENTS WHERE DISC='математика' AND MARK=5
```

Выведи список студентов, отсортировав его по убыванию алфавитного порядка

```
SELECT * FROM STUDENTS ORDER BY FAMILIA DESC
```

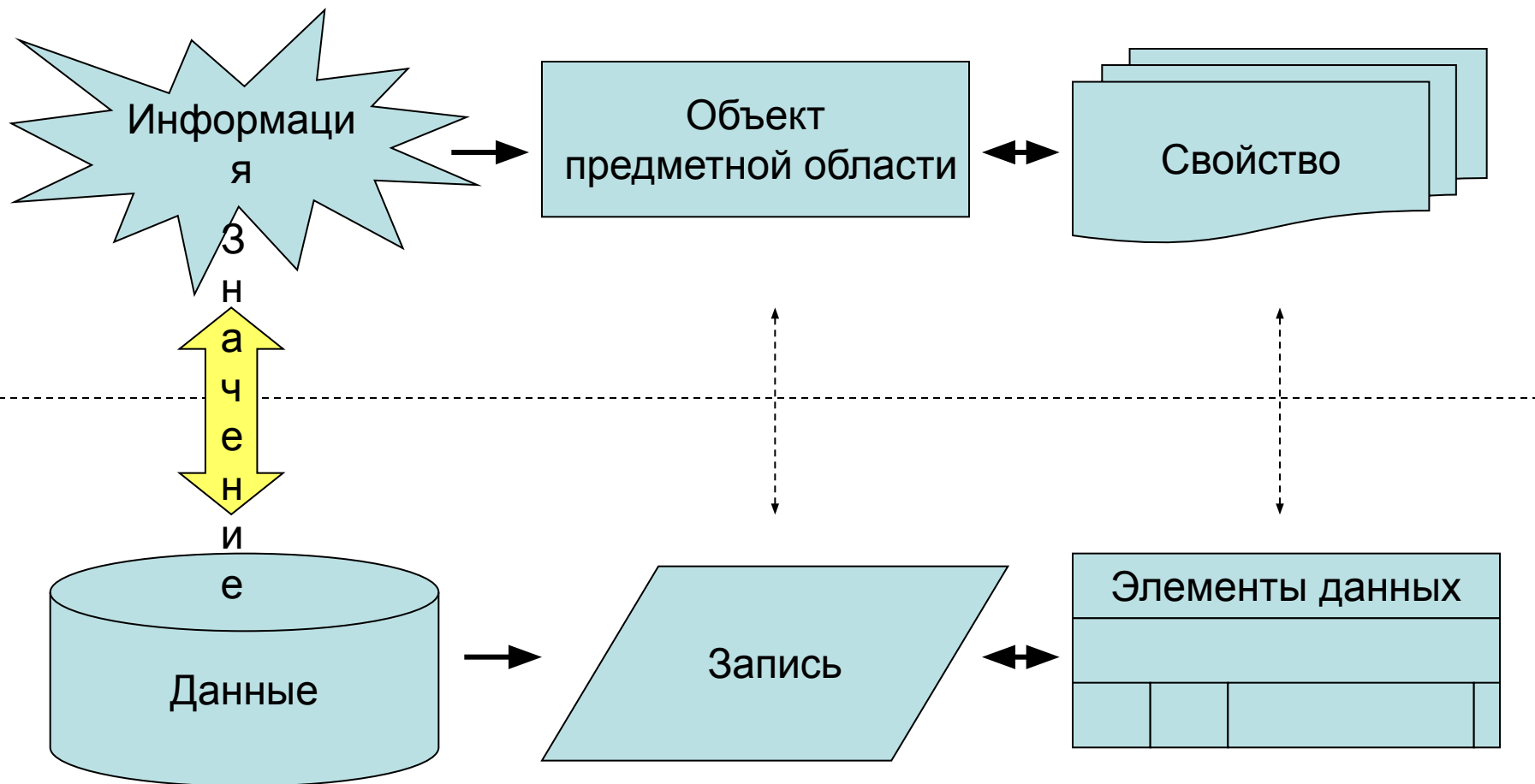
Выведи все оценки студента Петрова

```
SELECT * FROM OCENKI WHERE STUD_NOMER=(SELECT STUD_NOMER  
FROM STUDENTS WHERE FAMILIA='Петров')
```

Централизованное управление данными обеспечивает:

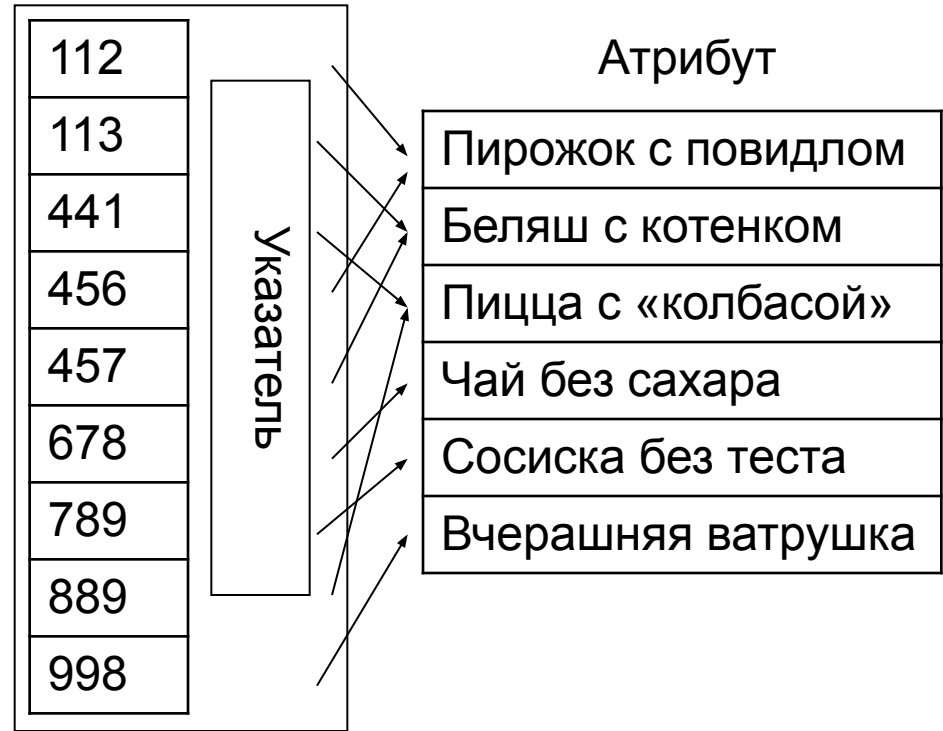
1. сокращение избыточности в хранимых данных;
2. совместное использование хранимых данных;
3. стандартизацию представления данных, упрощающую эксплуатацию БД;
4. разграничение доступа к данным;
5. целостность данных, обеспечиваемую процедурами, предотвращающими включение в БД неверных данных, и ее восстановление после отказов системы.

Атрибутивный способ идентификации



Способы хранения ключа и атрибута

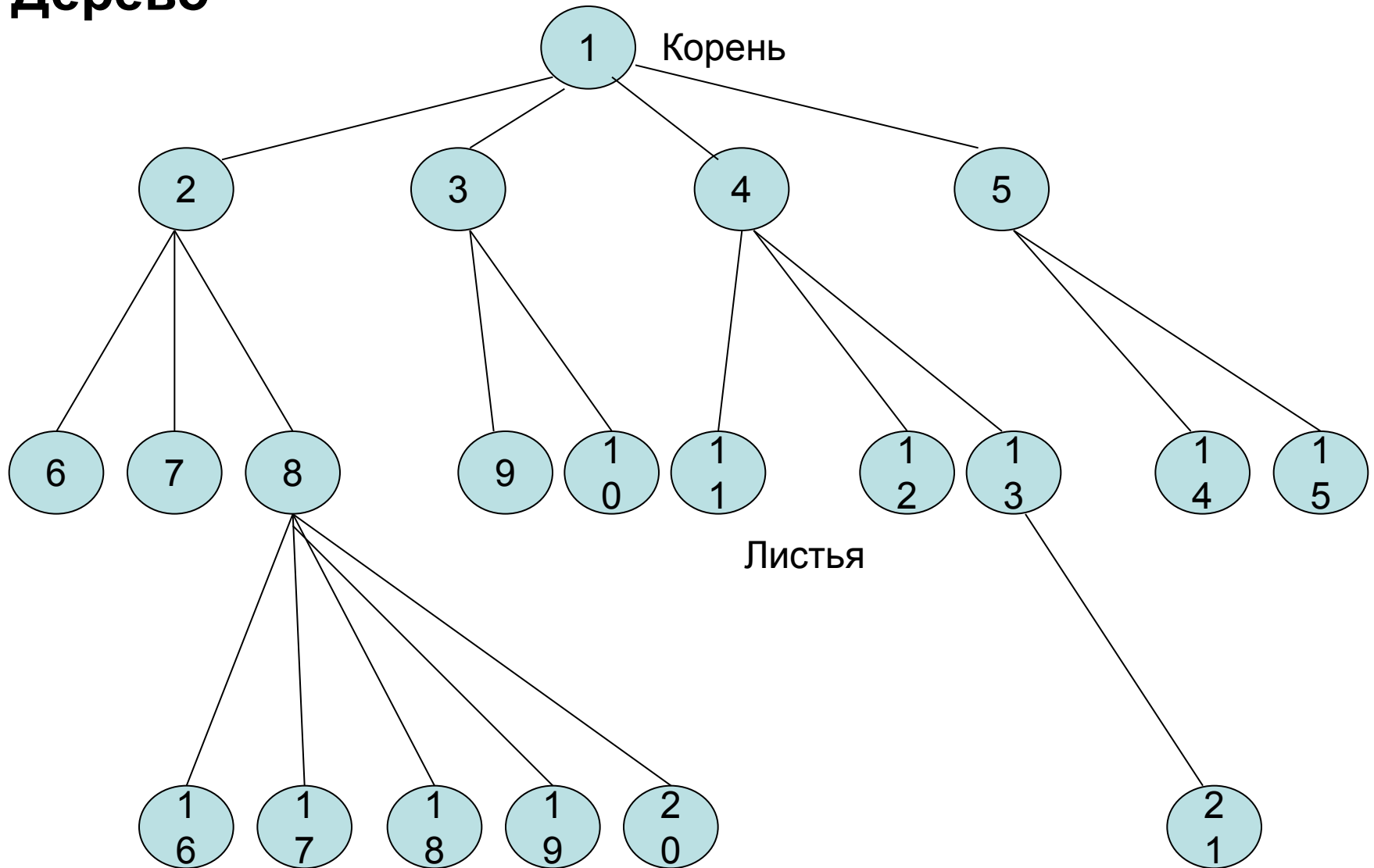
112	Пирожок с повидлом
113	Беляш с котенком
441	Пицца с «колбасой»
456	Пирожок с повидлом
457	Беляш с котенком
678	Чай без сахара
789	Сосиска без теста
889	Пицца с «колбасой»
998	Вчерашняя ватрушка



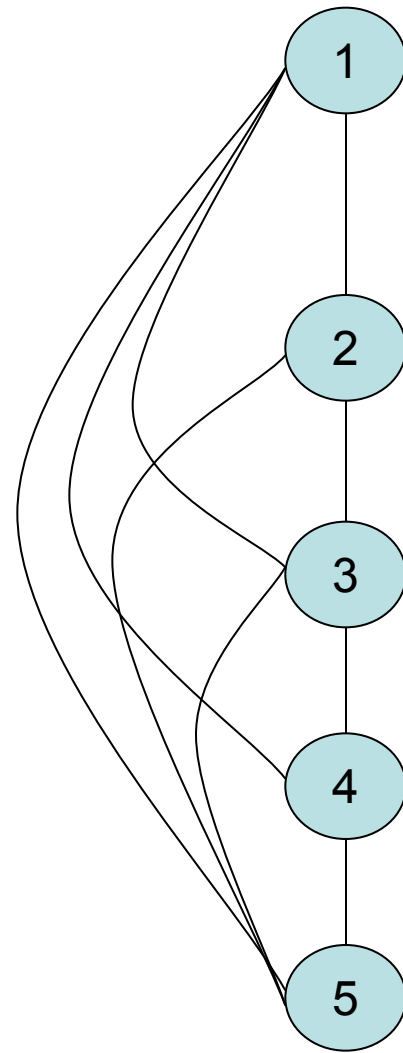
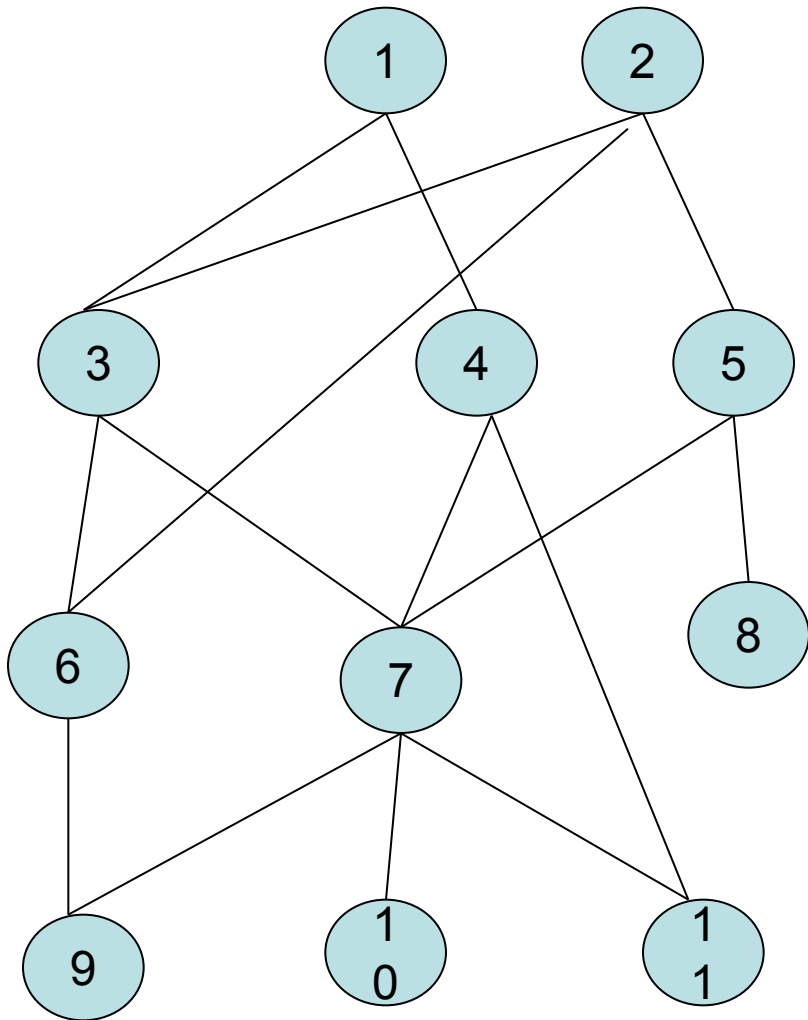
Инвертированный список

Пирожок с повидлом	112, 456
Беляш с котенком	113, 457
Пицца с «колбасой»	441, 889
Чай без сахара	678
Сосиска без теста	789
Вчерашняя ватрушка	998

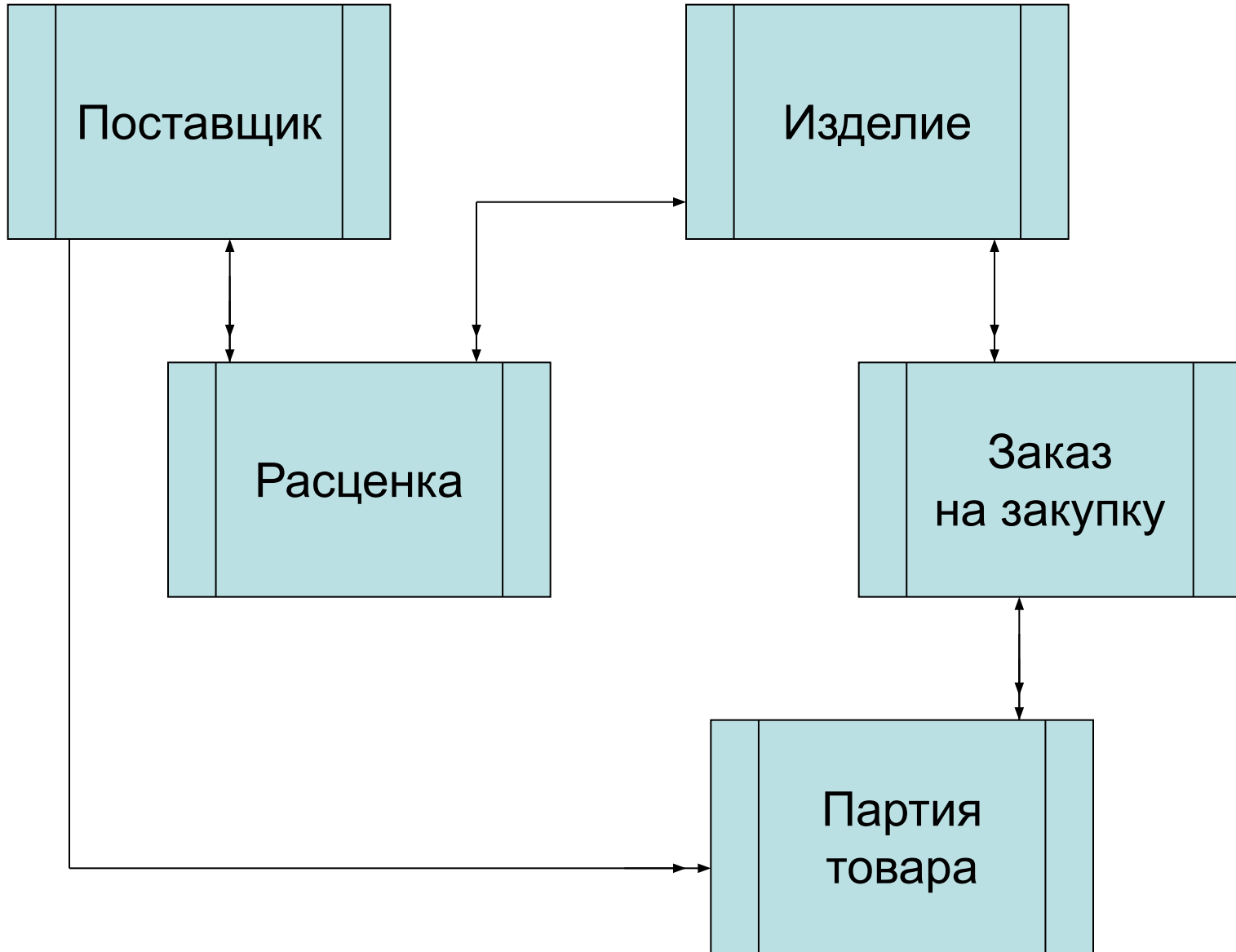
Дерево



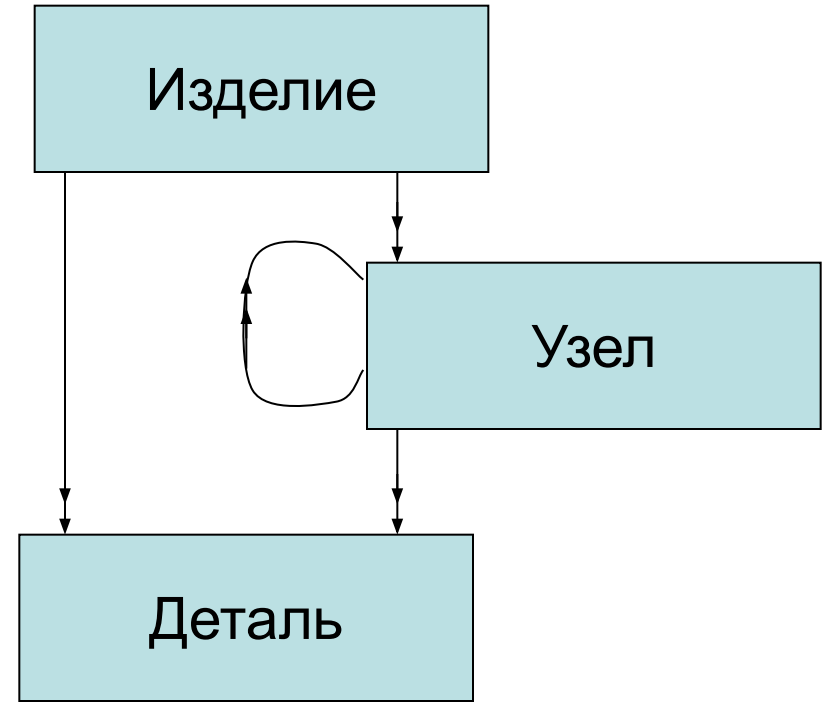
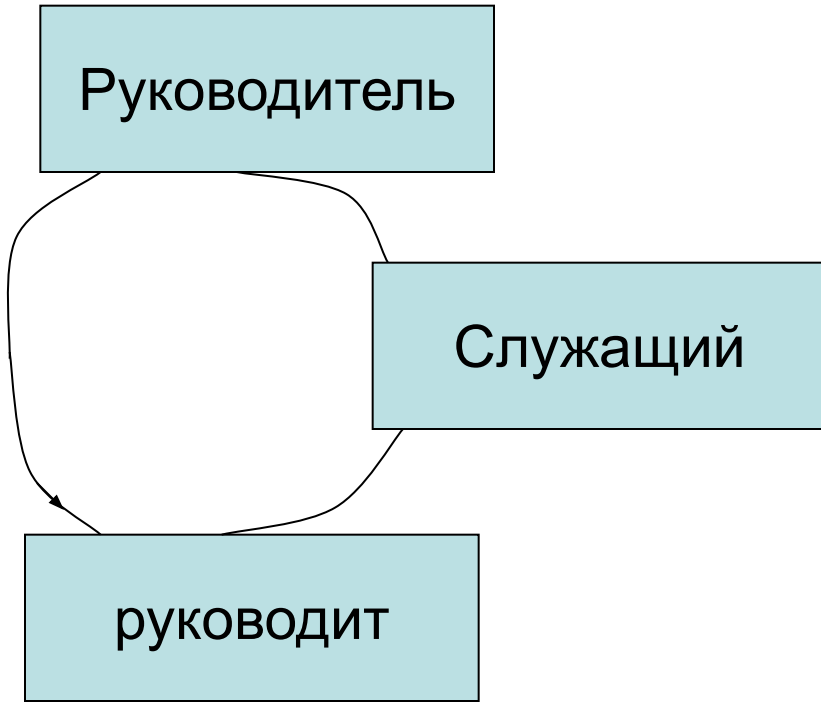
Сетевые структуры



Пример простой сетевой структуры



Пример сетевой структуры с петлей



Основные понятия реляционной модели



Реляционная модель предъявляет к таблицам следующие **требования**:

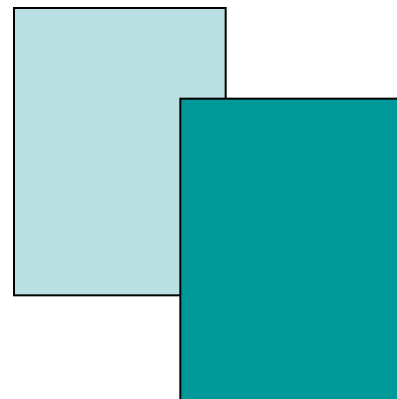
- 1) данные в ячейках таблицы должны быть структурно неделимыми ;
- 2) данные в одном столбце должны быть одного типа;
- 3) каждый столбец должен быть уникальным (недопустимо дублирование столбцов);
- 4) столбцы размещаются в произвольном порядке;
- 5) строки размещаются и таблице также в произвольном порядке;
- 6) столбцы имеют уникальные наименования.

Объединение

FIO	Year	Job	Kaf
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605

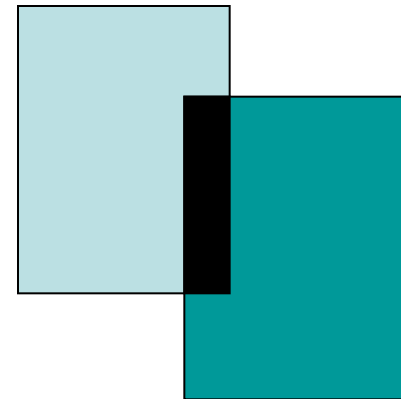


Пересечение

FIO	Year	Job	Kaf
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

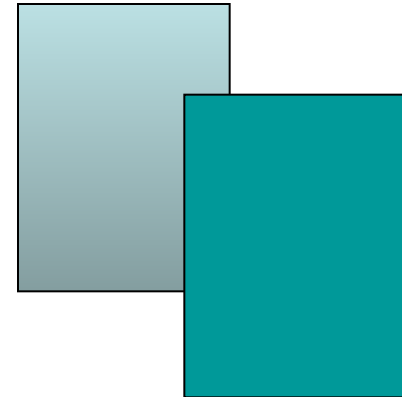


Разность

FIO	Year	Job	Kaf
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606

FIO	Year	Job	Kaf
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605



Произведение

Job
Профессор
Доцент
Ст.преподаватель
Ассистент

Kaf
606
605



Job	Kaf
Профессор	605
Профессор	606
Доцент	605
Доцент	606
Ст.преподаватель	605
Ст.преподаватель	606
Ассистент	605
Ассистент	606

Выборка

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605

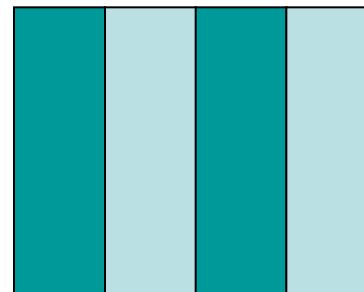
Kaf = 605



FIO	Year	Job	Kaf
Машкина М.М.	1978	Доцент	605
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605

Проекция

FIO	Year	Job	Kaf
Сидоров В.А.	1975	Профессор	606
Машкина М.М.	1978	Доцент	605
Дыркин Д.А.	1969	Ассистент	607
Петрова М.И.	1965	Ст.преподаватель	607
Сидорчук В.В.	1971	Ассистент	606
Пупкина Л.В.	1972	Доцент	605
Иванова И.И.	1974	Доцент	605



FIO	Job
Сидоров В.А.	Профессор
Машкина М.М.	Доцент
Дыркин Д.А.	Ассистент
Петрова М.И.	Ст.преподаватель
Сидорчук В.В.	Ассистент
Пупкина Л.В.	Доцент
Иванова И.И.	Доцент

Соединение

ФИО	Job
Сидоров В.А.	Профессор
Машкина М.М.	Доцент
Дыркин Д.А.	Ассистент
Петрова М.И.	Ст.преподаватель
Сидорчук В.В.	Ассистент
Пупкина Л.В.	Доцент
Иванова И.И.	Доцент

Job	ZP
Профессор	10000
Доцент	8000
Ст.преподаватель	5000
Ассистент	4000

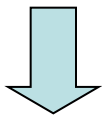
ФИО	Job	ZP
Сидоров В.А.	Профессор	10000
Машкина М.М.	Доцент	8000
Дыркин Д.А.	Ассистент	4000
Петрова М.И.	Ст.преподаватель	5000
Сидорчук В.В.	Ассистент	4000
Пупкина Л.В.	Доцент	8000
Иванова И.И.	Доцент	8000

Деление

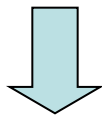
Делимое

Посредник

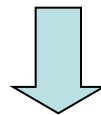
Делитель



Job
Профессор
Доцент
Ст.преподаватель
Ассистент



Job	Kaf
Профессор	606
Доцент	605
Ассистент	607
Ст.преподаватель	607
Ассистент	606
Доцент	605
Доцент	605



Kaf
606

=

Job
Профессор
Ассистент

Kaf
605
606

=

Job
Профессор
Доцент
Ассистент

Физическое представление с разделением данных и связей

Ф.И.О.	Год	Должность	Кафедра
Иванов И.И.	1948	Зав.кафедрой	22
Сидоров С.С.	1953	Профессор	22
Пупкин В.В.	1945	Профессор	22
Цветкова С.С.	1960	Доцент	22
Козлов К.К.	1959	Доцент	22
Петров П.П.	1960	Ст.преподаватель	22
Лютикова Л.Л.	1977	Ассистент	22
Сидорчук А.В.	1950	Зав.кафедрой	23
Пряников А.К.	1944	Профессор	23
Сумкин Ф.Б.	1958	Доцент	23
Соловьев Х.В.	1958	Доцент	23
Воробьева В.В.	1959	Ст.преподаватель	23
Дыркин В.С.	1966	Ассистент	23
Машкина М.М.	1976	Ассистент	23

1	Воробьева В.В.	1	1944	1	Ассистент
2	Дыркин В.С.	2	1945	2	Доцент
3	Иванов И.И.	3	1948	3	Зав.кафедрой
4	Козлов К.К.	4	1950	4	Профессор
5	Лютикова Л.Л.	5	1953	5	Ст.преподаватель
6	Машкина М.М.	6	1958		
7	Петров П.П.	7	1959	1	22
8	Пряников А.К.	8	1960	2	23
9	Пупкин В.В.	9	1966		
10	Сидоров С.С.	10	1976		
11	Сидорчук А.В.	11	1977		
12	Соловьев Х.В.				
13	Сумкин Ф.Б.				
14	Цветкова С.С.				

СВЯЗИ

Ф.И.О.	Год	Должность	Кафедра
3	3	3	1
10	5	4	1
9	2	4	1
14	8	2	1
4	7	2	1
7	8	5	1
5	11	1	1
11	4	3	2
8	1	4	2
13	6	2	2
12	6	2	2
1	7	5	2
2	9	1	2
6	10	1	2

Должность	Ф.И.О.	Год	Ф.И.О.	Кафедра	Ф.И.О.
1	5	1	3	1	3
1	2	2	9	1	10
1	6	3	3	1	9
2	14	4	11	1	14
2	4	5	10	1	4
2	13	6	12	1	7
2	12	6	13	1	5
3	3	7	1	2	11
3	11	7	4	2	8
4	10	8	7	2	13
4	9	8	14	2	12
4	8	9	2	2	1
5	7	10	6	2	2
5	1	11	5	2	6

Требования, предъявляемые к базам данных

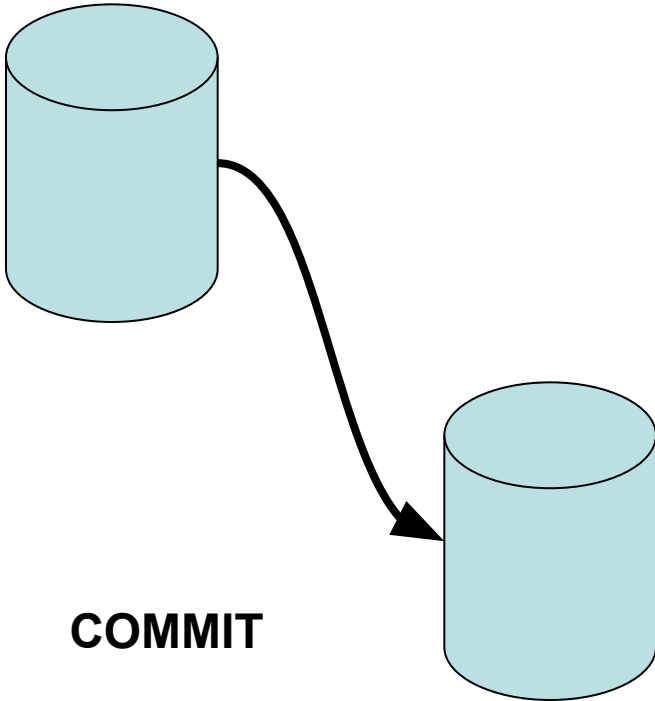
1. Описания должны быть понятны пользователю, не проектировавшему базу
2. Однажды принятые способы представления данных должны допускать присоединение новых элементов данных без изменения существующих схем данных и прикладных программ
3. СУБД должны позволять эффективно обрабатывать произвольные запросы к базе данных

Транзакции

Транзакция – неделимая с точки зрения воздействия на БД последовательность операторов манипулирования данными, такая, что:

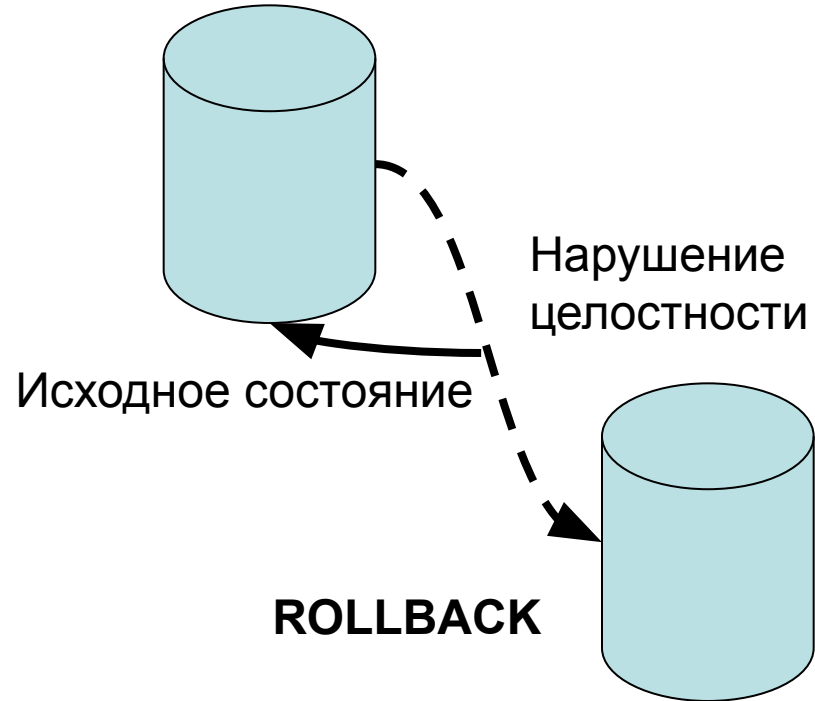
- 1) либо результаты всех операторов, входящих в транзакцию, отображаются в БД;
- 2) либо воздействие всех этих операторов полностью отсутствует.

Исходное состояние

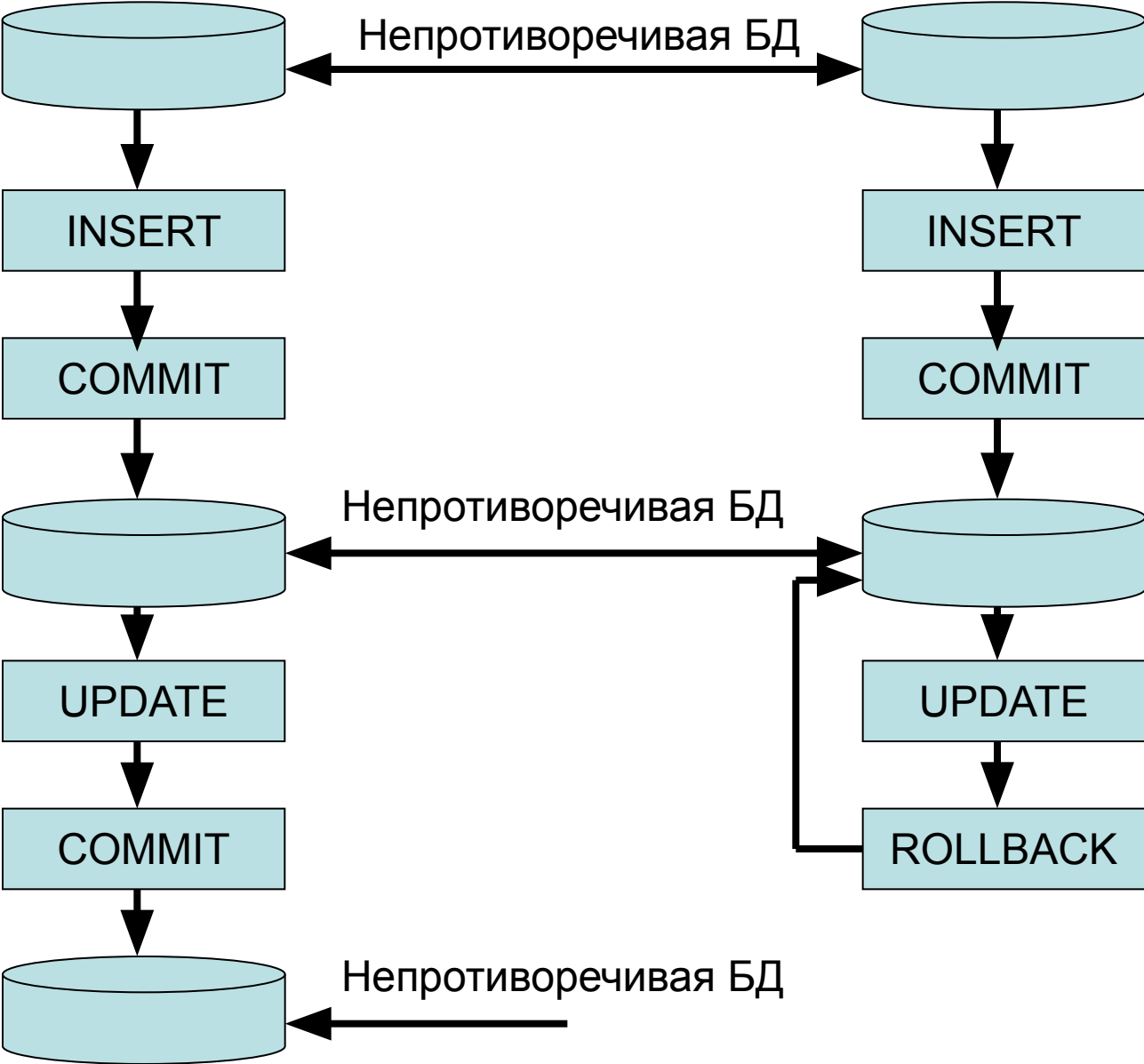


Измененная БД

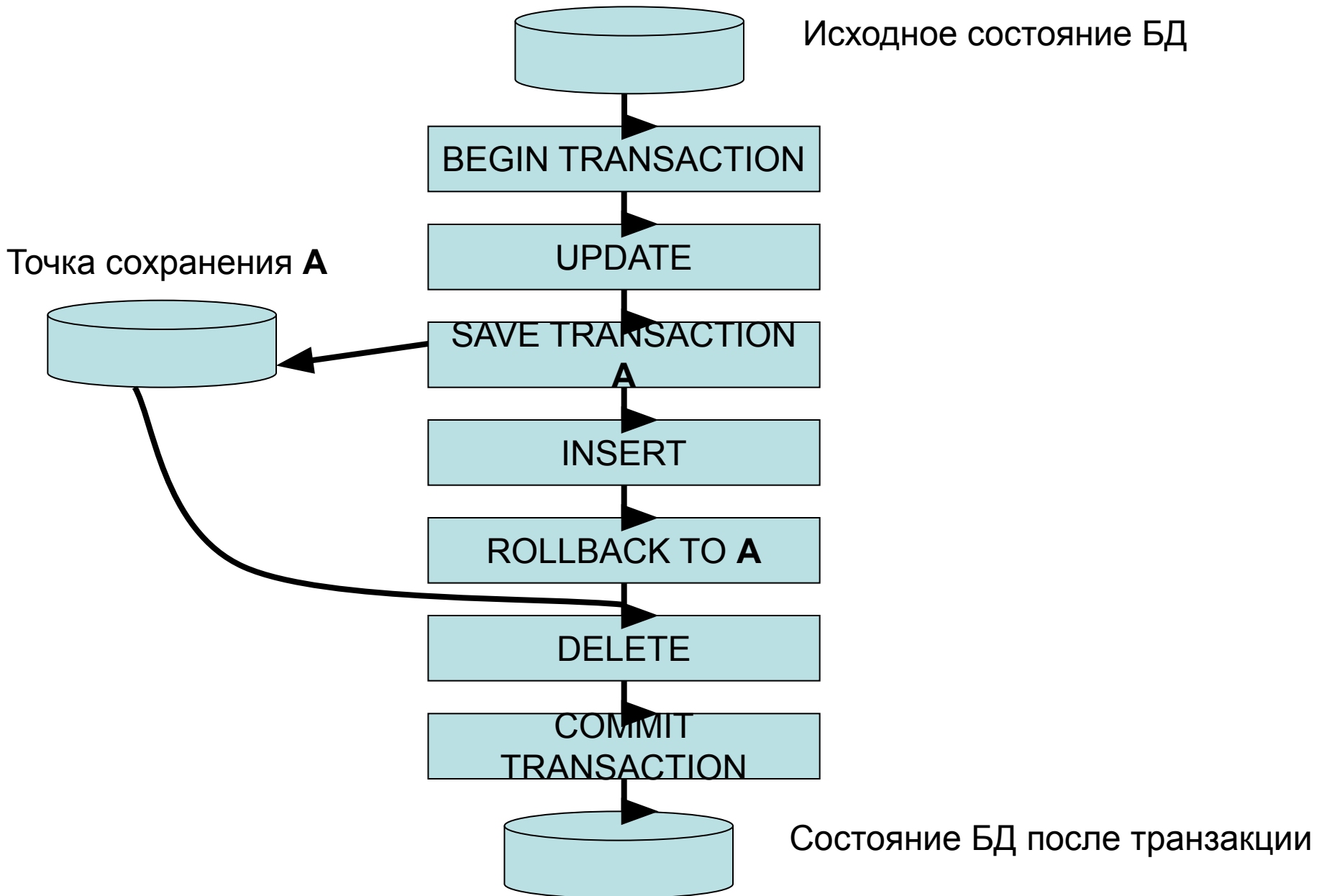
Исходное состояние



Модель автоматического выполнения транзакций



Модель управляемого выполнения транзакций



Журнал транзакций

Протокол журнализации (и управления буферизацией) действует по правилу Write Ahead Log (WAL) — «пиши сначала в журнал», и состоит в том, что если требуется сохранить во внешней памяти измененный объект базы данных, то перед этим нужно гарантировать сохранение во внешней памяти журнала записи о его изменении.

Общие требования к системе восстановления данных в составе СУБД

1. Пользователь не должен осуществлять рестарт транзакций или повторный ввод данных. Восстановление должно проходить на базе транзакции с помощью отмены или изменения отдельных транзакций.
2. Быстрое восстановление данных обеспечивается генерацией данных, используемых для восстановления.
3. При выполнении процедур автоматизированного восстановления пользователь не должен анализировать состав данных и выбирать сами процедуры.

Сериализация транзакций

Метод *сериализации транзакций* — это механизм их выполнения по такому плану, когда результат совместного выполнения транзакций эквивалентен результату некоторого последовательного выполнения этих же транзакций.

Между транзакциями могут существовать следующие виды конфликтов:

Транзакция 2 пытается изменять объект, измененный незакончившейся Транзакцией 1 (**W-W — конфликт**);

Транзакция 2 пытается изменять объект, прочитанный незакончившейся Транзакцией 1 (**R-W — конфликт**);

Транзакция 2 пытается читать объект, измененный незакончившейся Транзакцией 1 (**W-R — конфликт**).

Захват и освобождение объекта

Выделяются два основных режима захватов:

совместный режим — S (Shared), означающий разделяемый захват объекта и необходимый для выполнения операции чтения объекта;

монопольный режим — X (exclusive), означающий монопольный захват объекта и необходимый для выполнения операций записи, удаления и модификации.

Потенциально возможные объекты для захвата

В контексте реляционных баз данных возможны следующие варианты:

файл - физический (с точки зрения базы данных) объект, область хранения нескольких отношений и, возможно, индексов

таблица - логический объект, соответствующий множеству записей данного отношения;

страница данных - физический объект, хранящий записи одного или нескольких отношений, индексную или служебную информацию;

запись - элементарный физический объект базы данных.

Правила ACID

Транзакция — это законченный блок обращений к базе данных и некоторых действий над ней, для которого гарантируется выполнение четырех условий, так называемых свойств **ACID (Atomicity, Consistency, Isolation, Durability)**:

атомарность — операции транзакции образуют неразделимый атомарный блок с определенным началом и концом. Этот блок либо выполняется от начала до конца, либо не выполняется вообще. Если в процессе выполнения транзакции произошел сбой, происходит откат к исходному состоянию

согласованность — по завершении транзакции все задействованные объекты находятся в согласованном состоянии

изолированность — одновременный доступ транзакций различных приложений к разделяемым объектам координируется таким образом, чтобы эти транзакции не влияли друг на друга

долговременность — все изменения данных, осуществленные в процессе выполнения транзакции, не могут быть потеряны