

БАНКИ ДАННЫХ

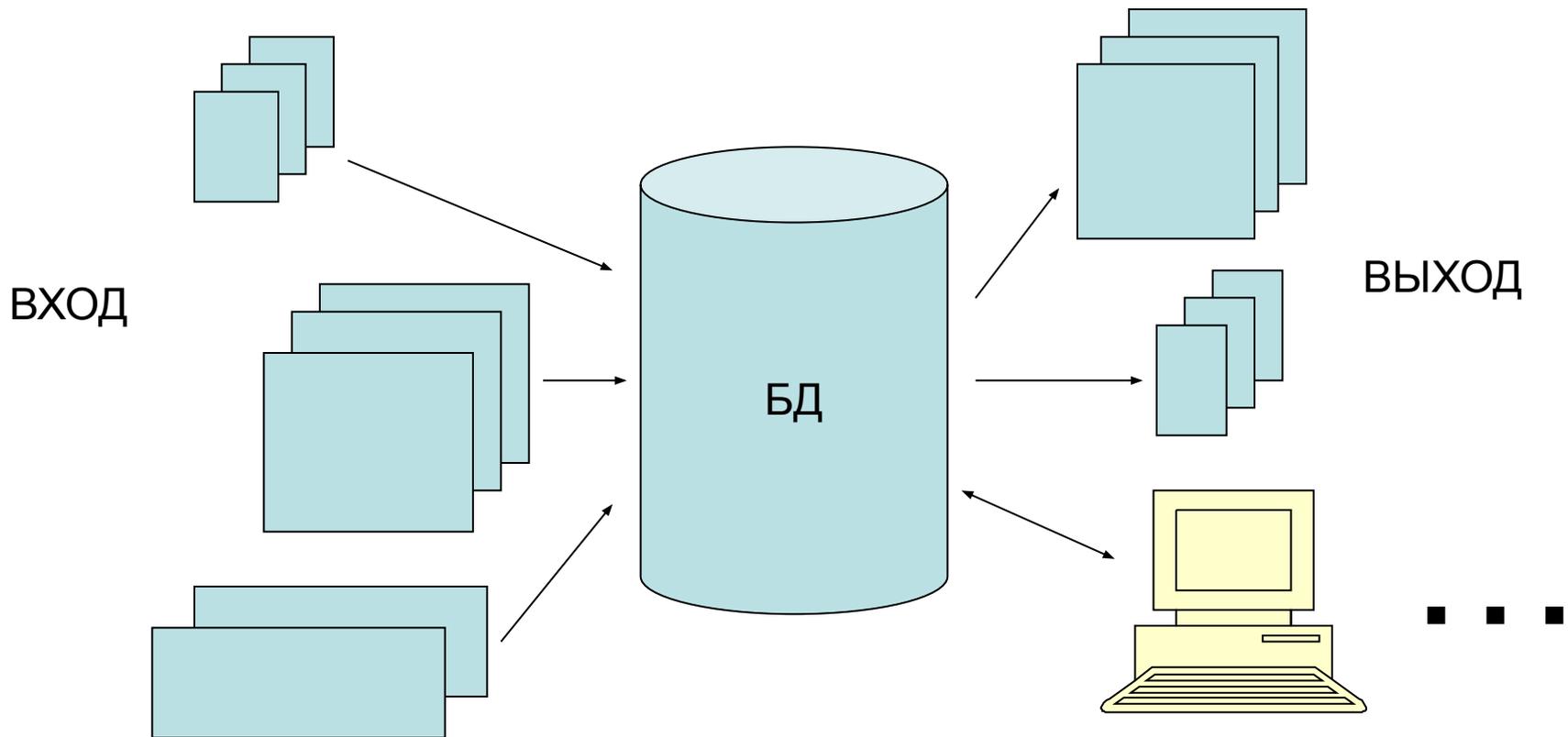
Автор: Емельянов Н. Е.
Правка: Тригуб Н.А.

6. Интерфейс с БД

Проблемы создания БД

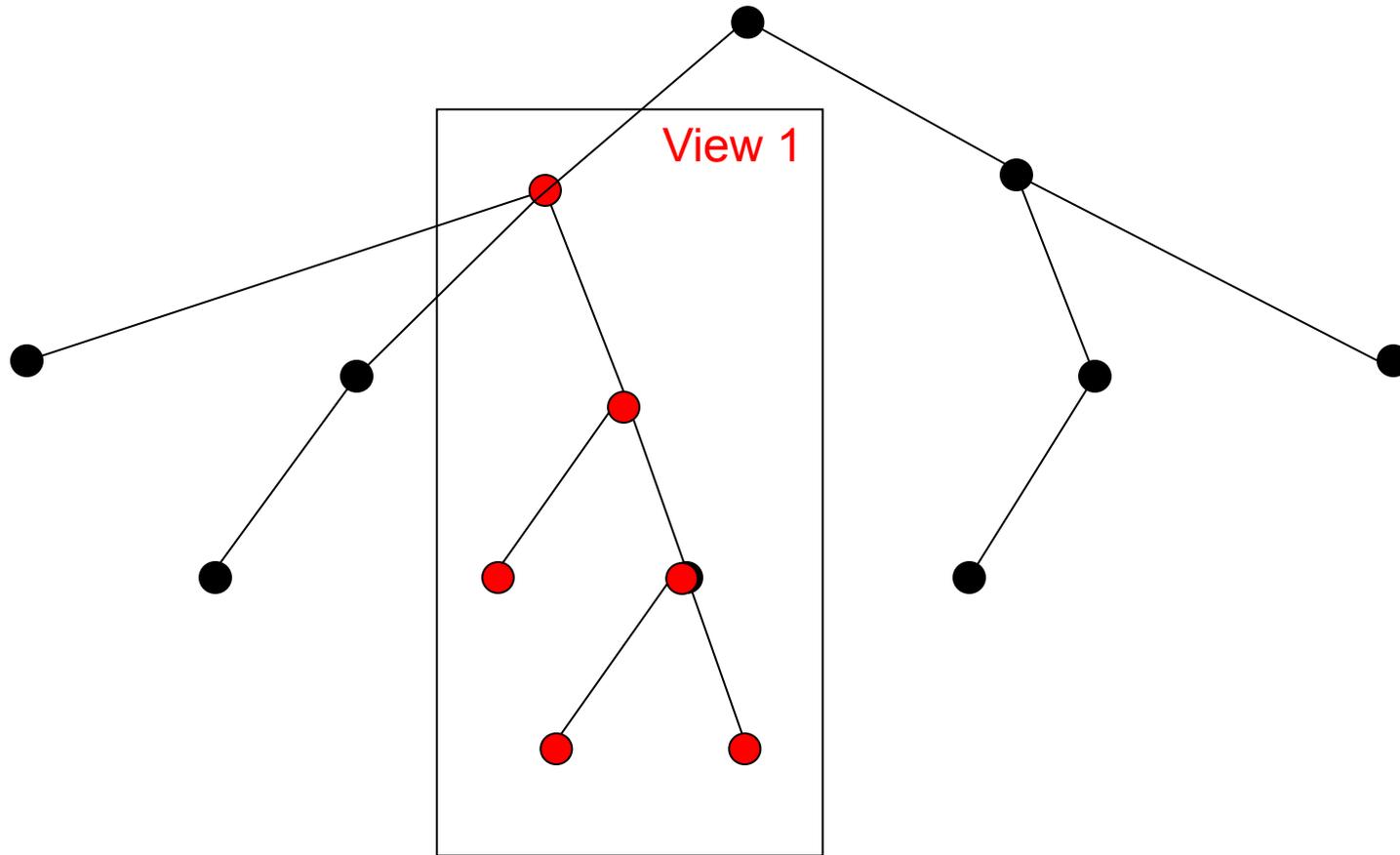
- Проектирование (концептуальное)
- Хранение (выбор СУБД, логическое проектирование)
- Интерфейс (как организовать ввод-вывод)
- Технология эксплуатации (сеть или нет и др.)

6.1. Понятие документа

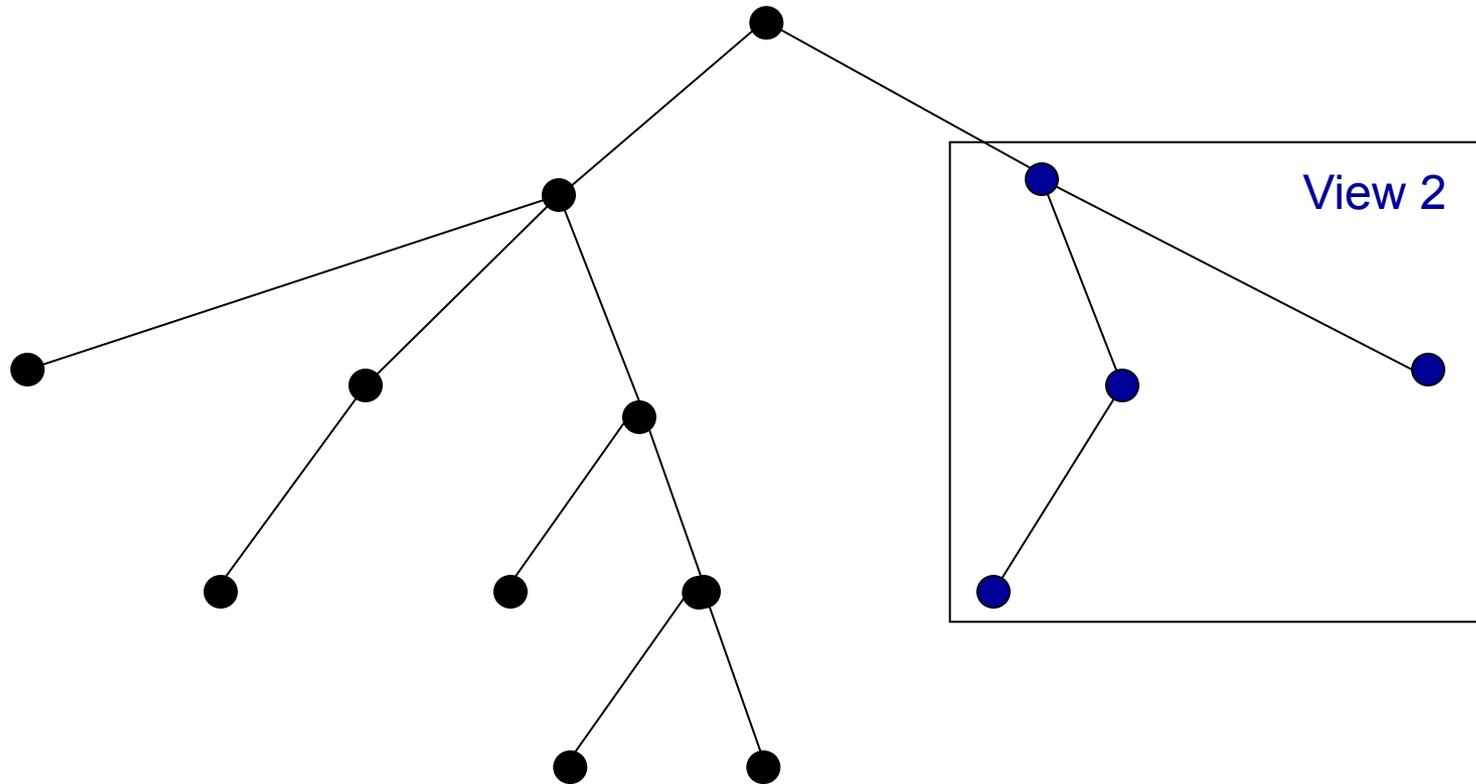


Малый бизнес – до 100 дисплеев
Средний - от 100 до 1000
Большой – более 1000

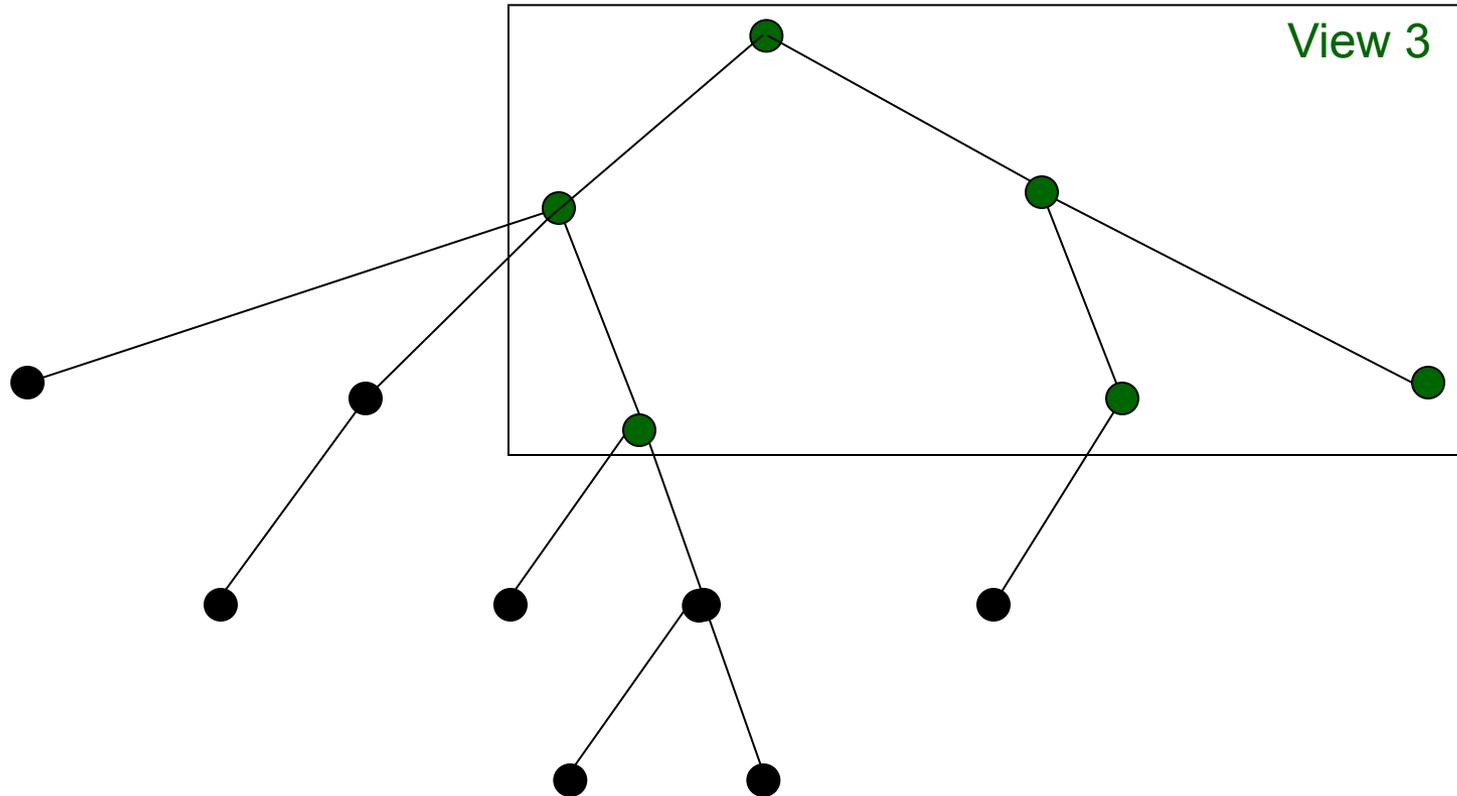
Понятие User View - взгляд пользователя



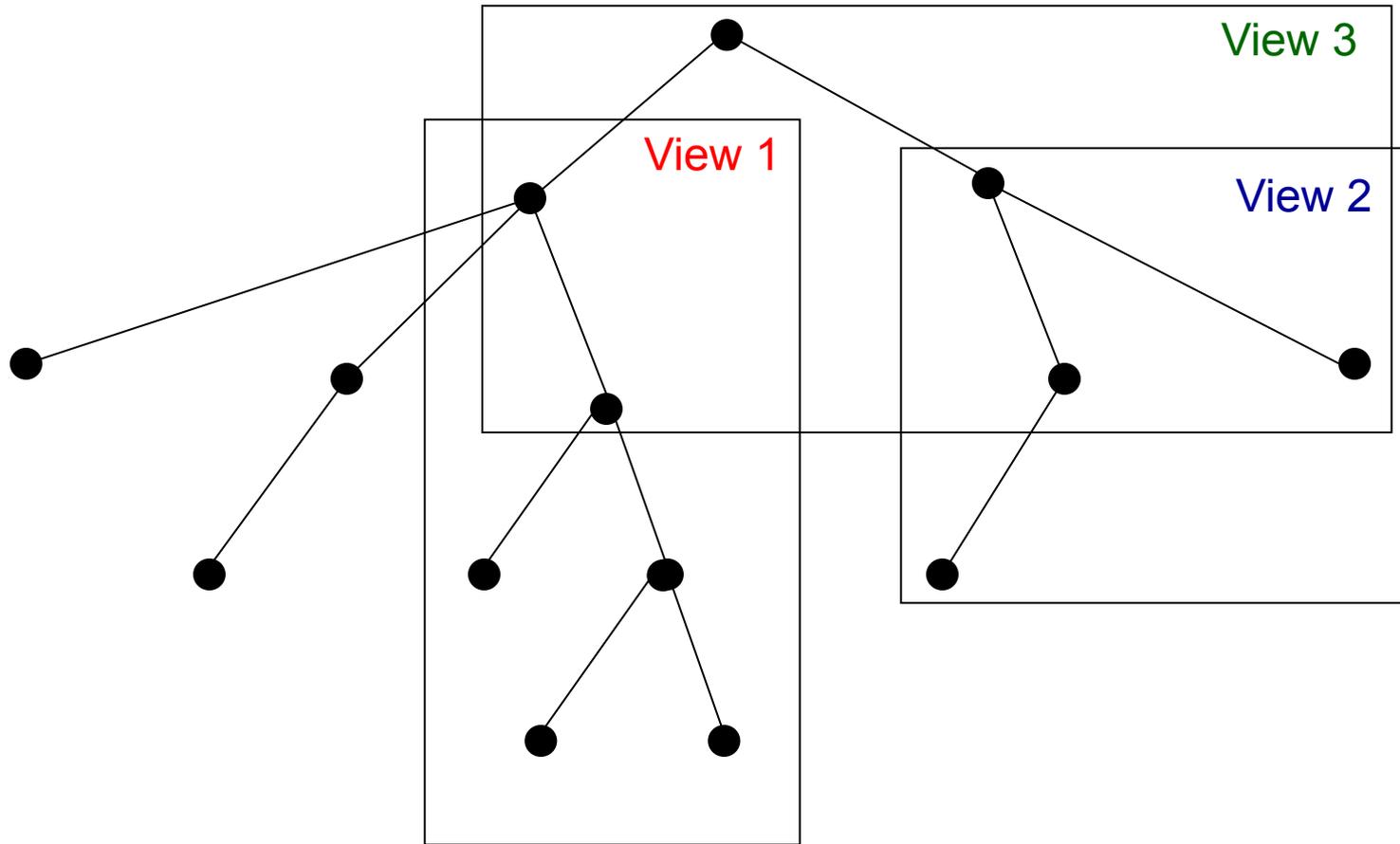
Понятие User View - взгляд пользователя



Понятие User View - взгляд пользователя



Понятие User View - ВЗГЛЯД ПОЛЬЗОВАТЕЛЯ



Каждый документ – один из
взглядов (View).

Сколько же документов в ИС?

	Среднее	МАХ
Вход	12	200
Выход	20	300

Обследование 5.000 ИС в 2000 г.

***Документ* (D) –**
структурированный текст
(справка, таблица, газета, книга и
т.д.) на магнитном носителе или в
последовательности символов.

$$D = \bigcup_i V_i$$

Электронный документ –
структурированная информация
для ввода, вывода, передачи,
согласования, подписи и т.д. на
магнитном носителе или в
последовательности символов.

6.2. Форма и содержание документа

АНКЕТА	
ФИО	<input type="text" value="Иванов И.И"/>
Г.Р.	<input type="text" value="1986"/>
.....	

Форма – постоянная часть всех документов данной форма.

Содержание – переменная часть.

Министерство высшего образования

МИСИС

Кафедра ИК

Фонд заработной платы

Январь 2008 года

500.000 рублей

Министерство высшего образования

МИСИС

Кафедра ИК

Фонд заработной платы

Январь 2008 года

500.000 рублей

Если ИС – кафедральная зарплата

Министерство высшего образования
МИСИС

Кафедра ИК

Фонд заработной платы

Январь 2008 года
500.000 рублей

Если ИС – зарплата МИСИС

Министерство высшего образования

МИСИС

Кафедра ИК

Фонд заработной платы

Январь 2008 года
500.000 рублей

Если ИС – Министерства высшего образования

Министерство высшего образования

МИСИС

Кафедра ИК

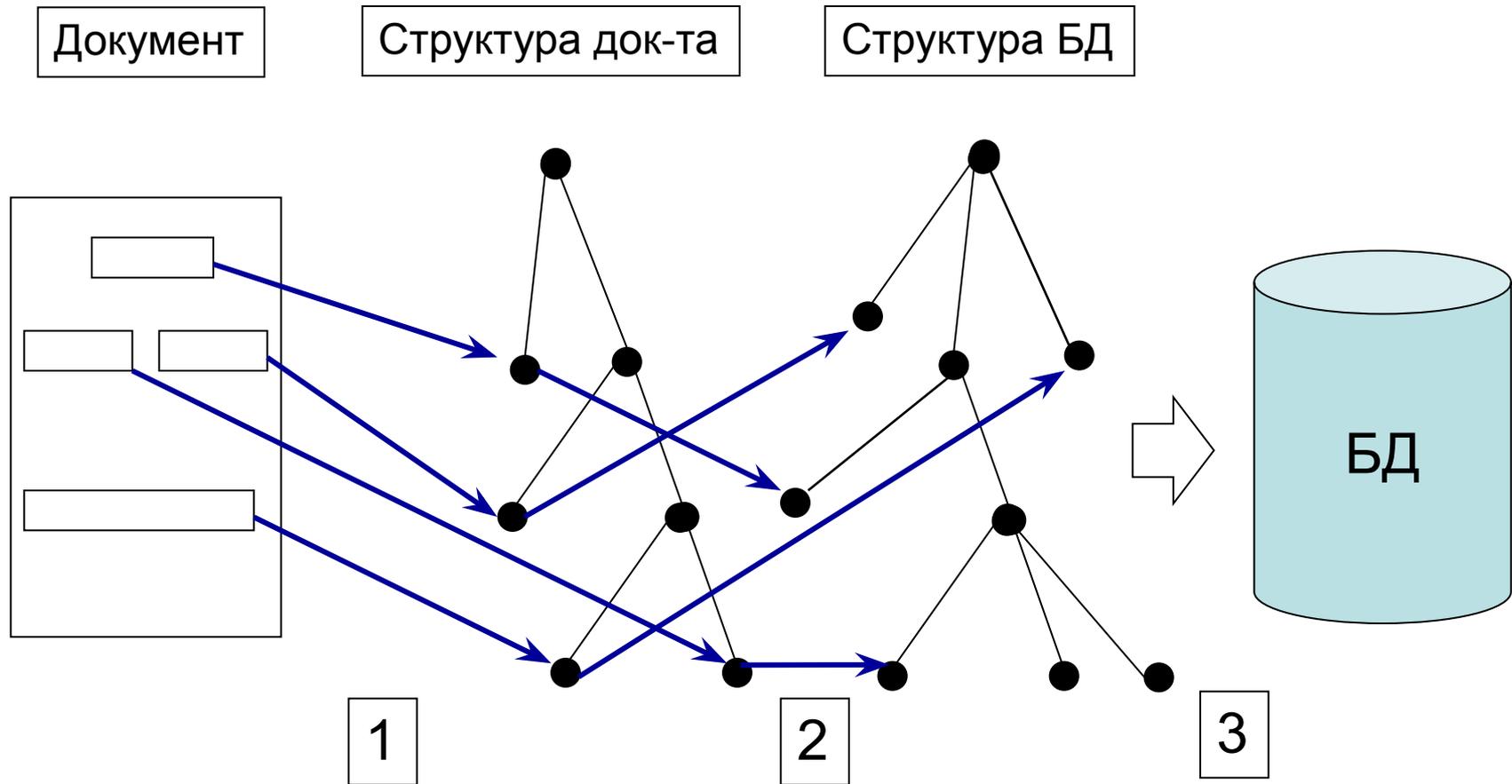
Фонд заработной платы

Январь 2008 года

500.000 рублей

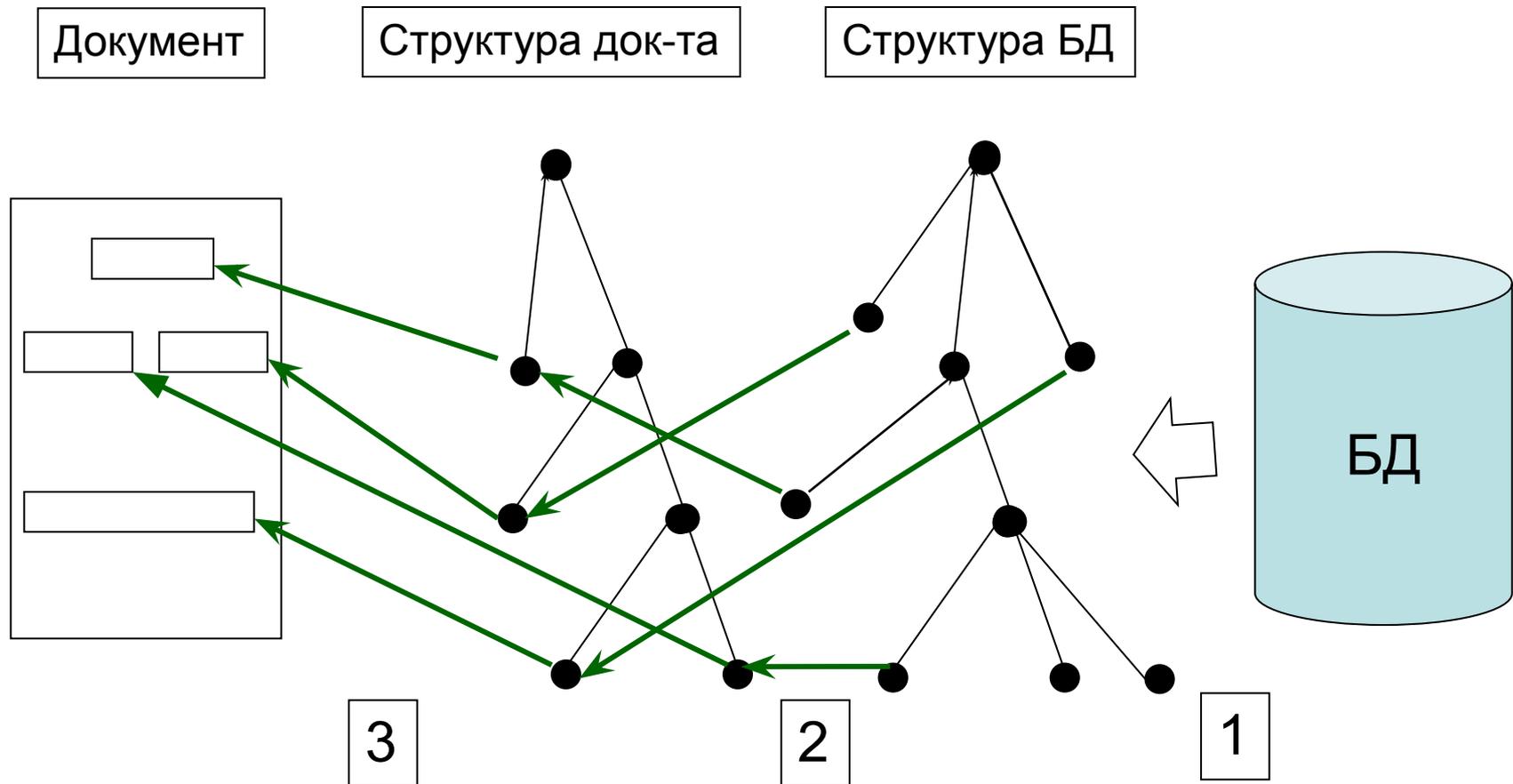
Если ИС – Совета Министров

6.3. Ввод документов



- 1 – расформатирование,
- 2 – реструктуризация (из док. в БД),
- 3 – запись в БД

6.4. Вывод документов



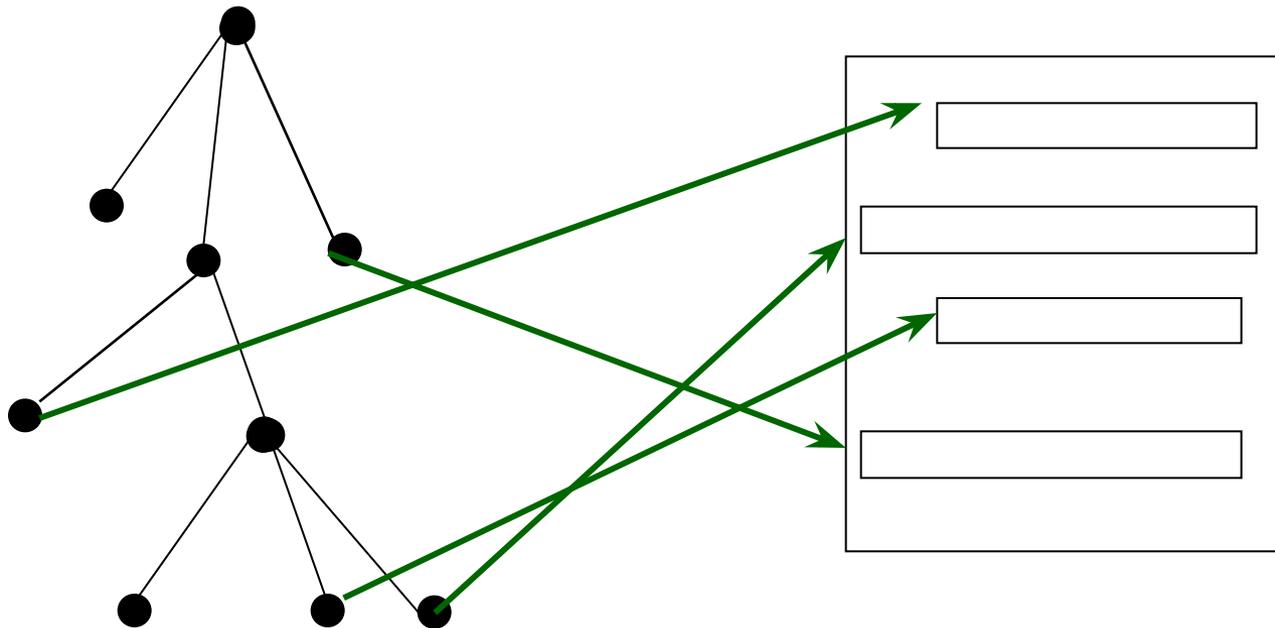
- 1 – отбор из БД,
- 2 – реструктуризация (из БД в док.),
- 3 – форматирование

Вопрос. Можно ли повторно вводить в БД один и тот же документ?

- $F(\text{Б (старая)}, D_i^j) \Rightarrow \text{Б (нов.)}$
- $F((\text{Б (ст.)}, D_i^j), D_i^j) \Rightarrow \text{Б' (нов.)}$

-
- $\text{Б (нов.)} = \text{Б' (нов.)} ?$
 - Может ли быть, что
 $\text{Б (нов.)} \neq \text{Б' (нов.)} ?$

6.5. Классификация видов представления данных в документах



Как можно отобразить сложные структуры на лист бумаги ?

- *Аксиома 1.* Не бывает документов типа **Зарплата = 5000.**
Обязательно есть идентификаторы, которые определяют данные.
- *Аксиома 2.* Семантические блоки или не пересекаются или один входит в другой.

То есть кроссворды

М
О
А С У
К
В
А

не документы.

Если Аксиомы 1, 2 выполнены, то можно все отображения разбить на следующие типы.

Типы отображений

1) Последовательности

- - с именами
- - с разделителями
- - семантические

2) Таблицы

- - с промежуточными заголовками
- - с заголовками столбцов
- - с заголовками строк

3) Иерархии

- - с номерами уровней
- - отступами
- - контурами
- -ссылками

4) Смешанное представление

Типы отображений

1) Последовательности

- - с именами Вес = 150 (имя = знач.)
- - с разделителями (3 x 10 см.) (размеры)
- - семантические Москва, 2008 (город и год)

2) Таблицы

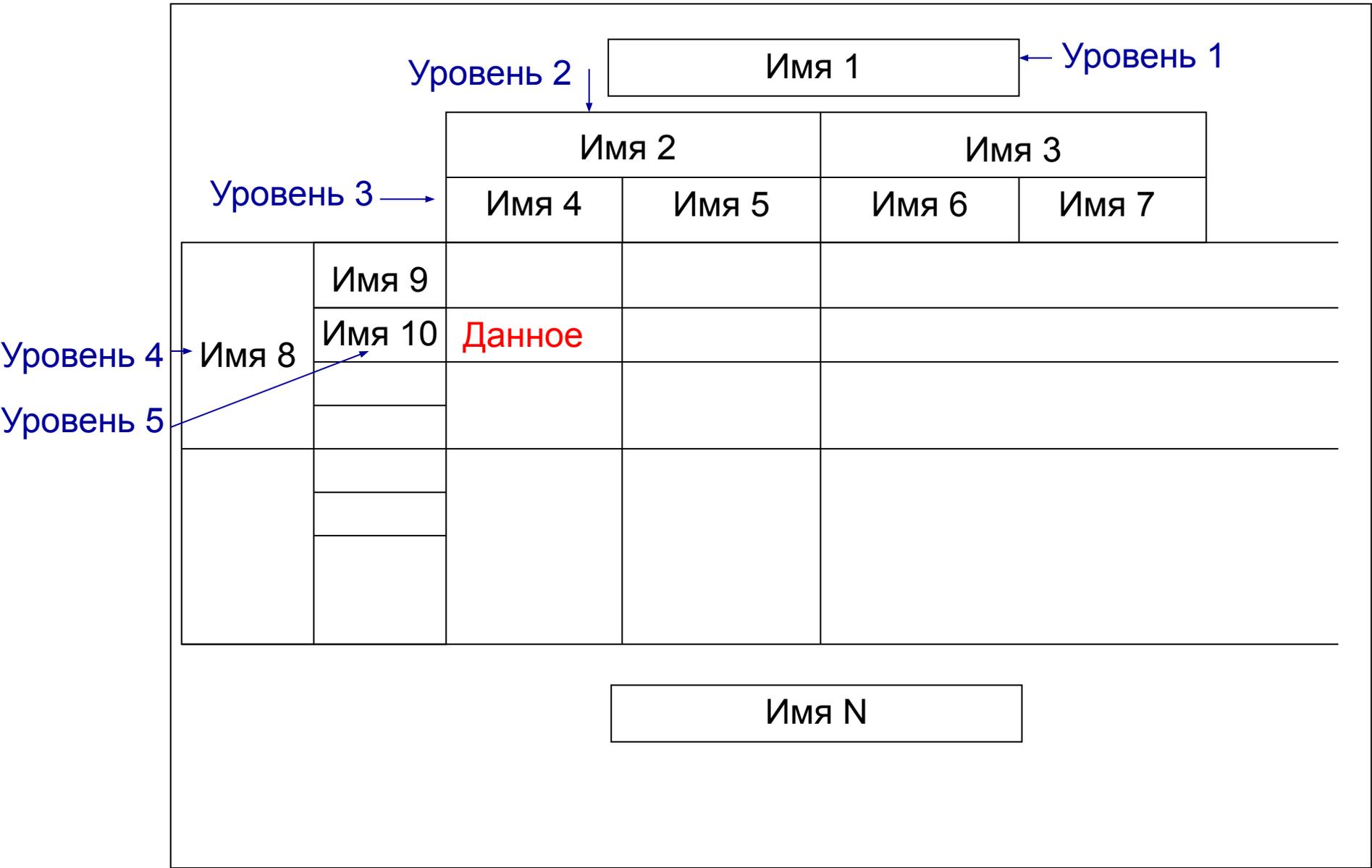
- - с промежуточными заголовками
- - с заголовками столбцов
- - с заголовками строк

3) Иерархии

- - с номерами уровней
- - отступами
- - контурами
- -ссылками

4) Смешанное представление

Таблицы



Иерархии

Номера уровней

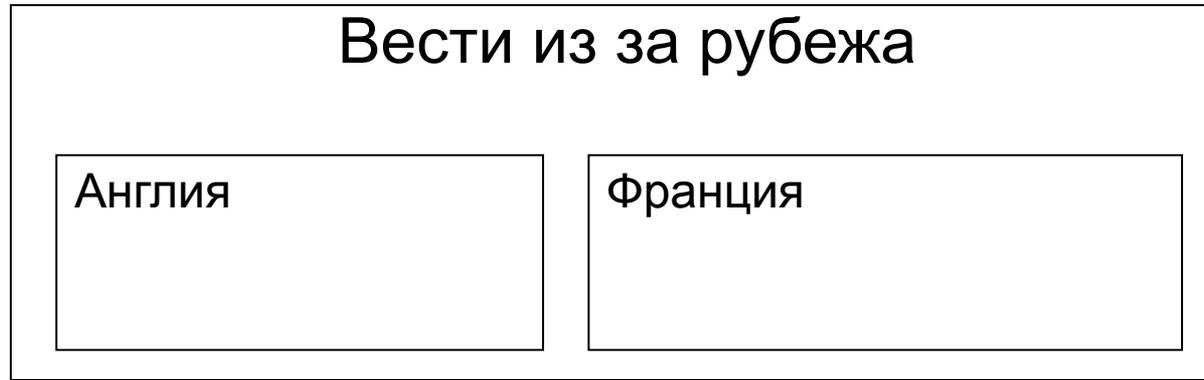
- 1.
- 1.1.
- 1.2.
- 1.2.1.

Отступы (блок – схемы)

- A
 - B
 - C
 - └─ D

Иерархии

- Контурь (газетное представление)

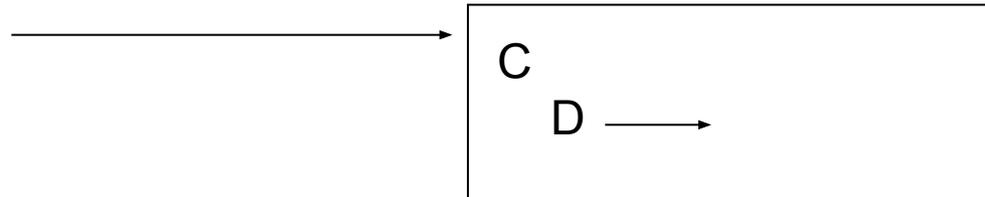


- Ссылки

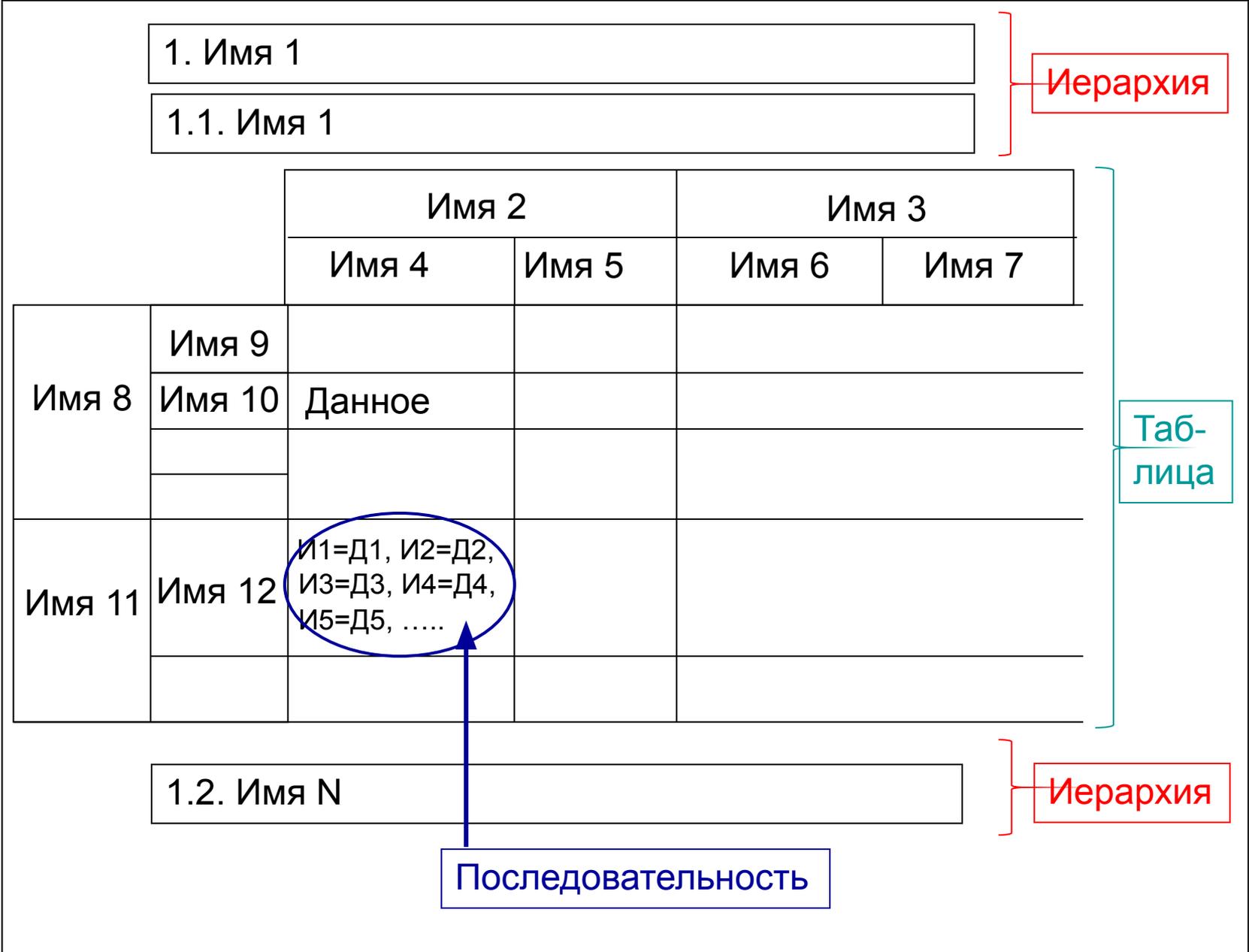
- А

- В →

- С



Смешанное представление

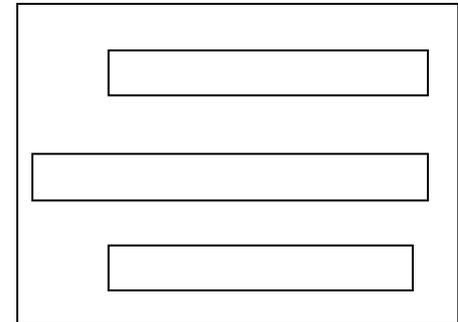


Достоинства типов представления

	Компакт- ность	Инверсия	Сложные структуры
Последова- тельности (каталоги)	+		
Таблицы (бизнес)		+	
Иерархии (родослов- ные)			+

6.6. Современный электронный документ (Система Lotus)

Документ – структурированный текст



Приложения:

ИЗО

AUDIO

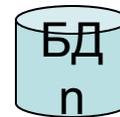
VIDEO

Тексты

Базы данных:



...



Электронная цифровая подпись

7. Запрос к БД

7.1. Компоненты запроса

- 1) Что выбрать
 - 2) По каким условиям
 - 3) Как сортировать
 - 4) Как форматировать
-
- 5) И 2) – функции языка запросов
 - 3) И 4) – генератора отчетов

Три способа построения запросов

- Алгебра
- Исчисление
- Языки запросов

7.2. Реляционная алгебра Кодда

- Всего 6 одно и двухместных операторов

$$F(A) \rightarrow B$$

$$G(A, B) \rightarrow C$$

где A, B, C – отношения

Кодд доказал, что 6 операторов достаточно для выполнения любого запроса

Операторы реляционной алгебры

1. Объединение (Union)

$$U(A, B) = C$$

A и B одинаковы по составу реквизитов.

К строкам A добавляются строки B

2. Пересечение (Intersection)

$$I(A, B) = C$$

A и B одинаковы. Выделение общих строк.

3. Вычитание (Difference)

$$D(A, B) = C$$

A и B одинаковы. Вычитаются из A общие с B строки.

4. Выбор (Select)

$$S(A, \text{условие}) = B$$

Выбор из A строк, удовлетворяющих условию.

5. Проекция (Project)

$$P(A, \text{имена столбцов}) = B$$

Выбор из A указанных столбцов.

6. Соединение (Join)

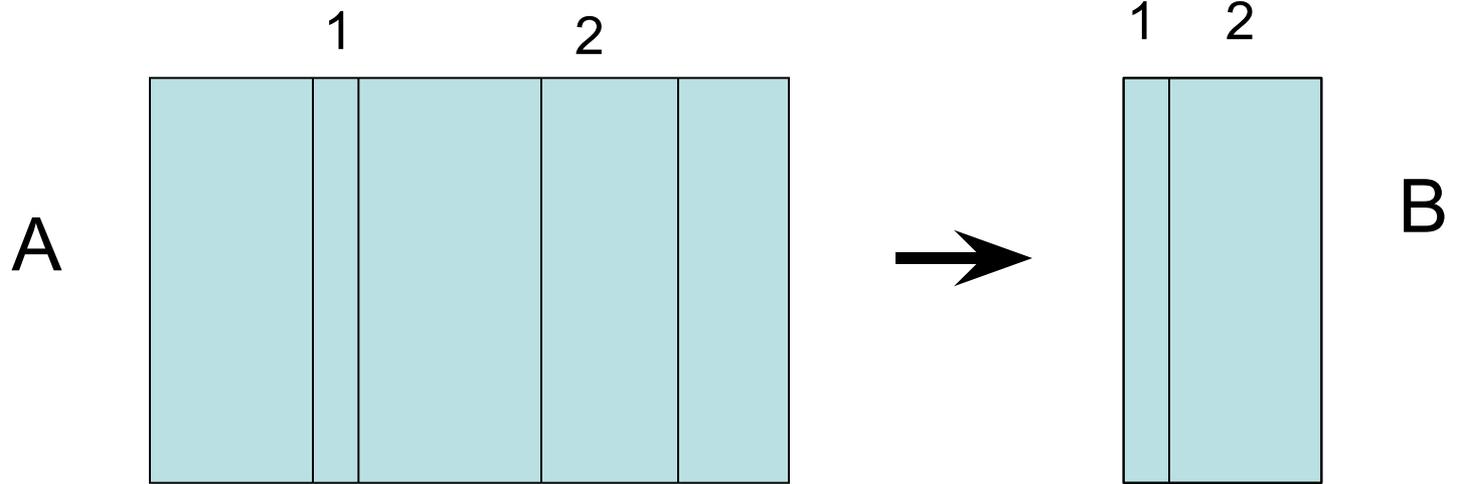
$$J(A, B, \text{условие}) = C$$

A и B разного состава. Объединение строк A и B.

$$U(A,B) = C$$

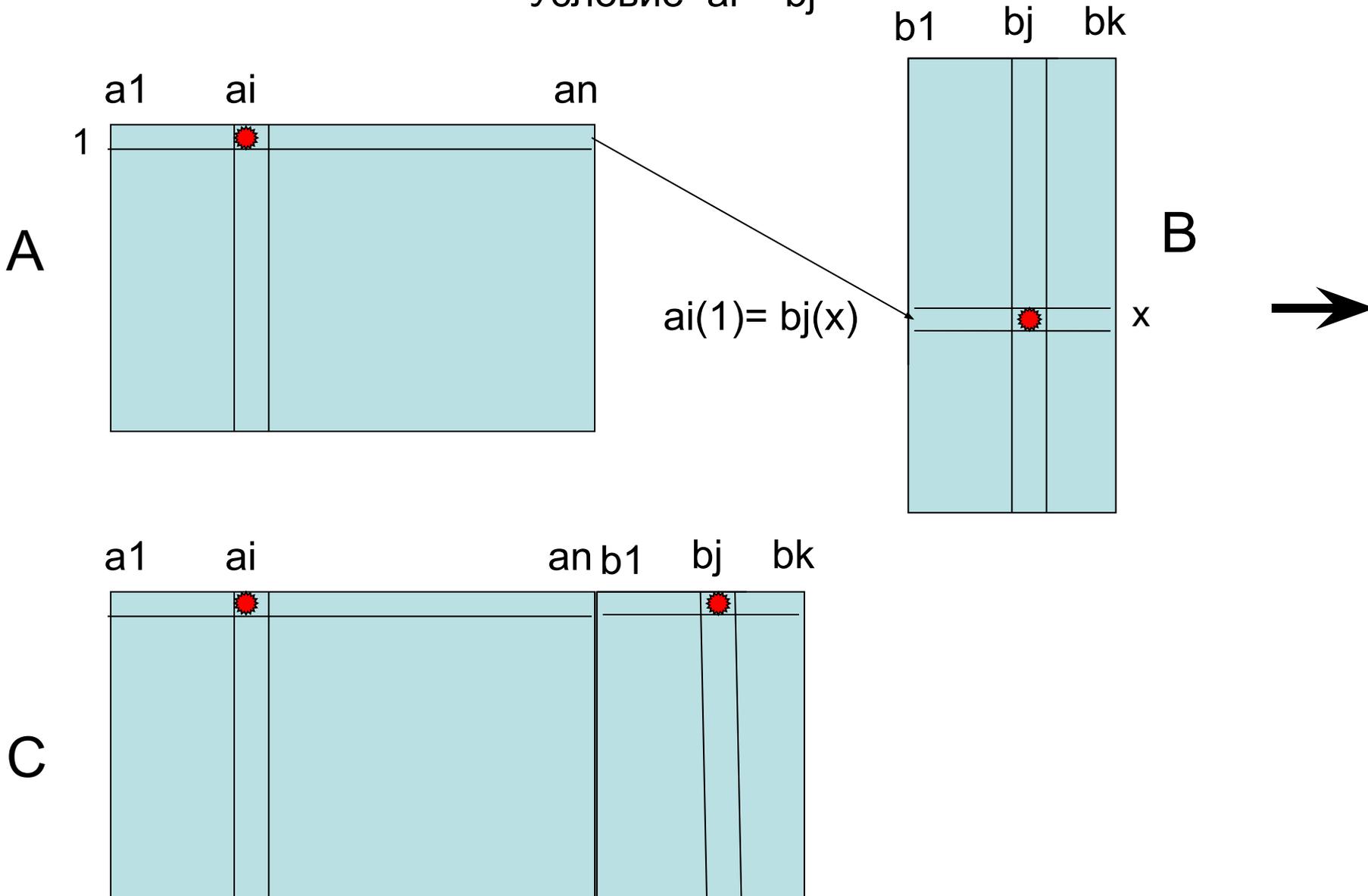


$$P(A, \text{имена столбцов}) = B$$



$$J(A, B, \text{условие}) = C$$

Условие $a_i = b_j$



7.3. Реляционное исчисление (calculus) Кодда

- Исчисление – знаковая система из термов (слов), формул (фраз), правил вывода новых правильных формул из уже имеющихся

- Знаки

$\{ \}$ – отношение

$A(a,b,\dots)$ – имя и реквизиты отношения

$:$ — справа стоит условие

Пример

- $\{Sp(p\#, s\#)\} \leftarrow \{S(City, s\#)\}$
- $\{C(Sp.p\#, S.s\#) : Sp.s\# = S.s\#\}$
- Новое отношение C построено при помощи Join и Project

7.4. Языки запросов

7.4.1. SQL (Structured Query Language)

- Структурированный ЯЗ.
Стандарты ISO 1989 (1992)
- Операторы
Select – что выбрать
From – откуда выбрать
Where – при каких условиях

Примеры запросов

Результат

a) S (s#, city, status) - поставщики
Select s#, status
From S
Where city = 'Paris' and status > 20

Поставщики
Парижа

S#	Status
----	--------

b) S - то же
Sp (s#, p#) - поставки
Select p#, city
From Sp, S
Where Sp.s# = S.s#

Список всех
деталей и городов
их производства

p#	city
----	------

c) S, Sp - те же
P (p#, color) - детали
Select s#, city
From S
Where s# In
 Select s#
 From Sp
 Where p# In
 Select p#
 From P
 Where color = 'red'

Поставщики
красных деталей

S#	City
----	------

- Алгебра описывает **как** выполняется запрос
- Исчисление и Язык запросов определяют **что** нужно получить

7.4. Языки запросов

7.4.2. XQuery (XML - запрос)

- Для XML - DB.

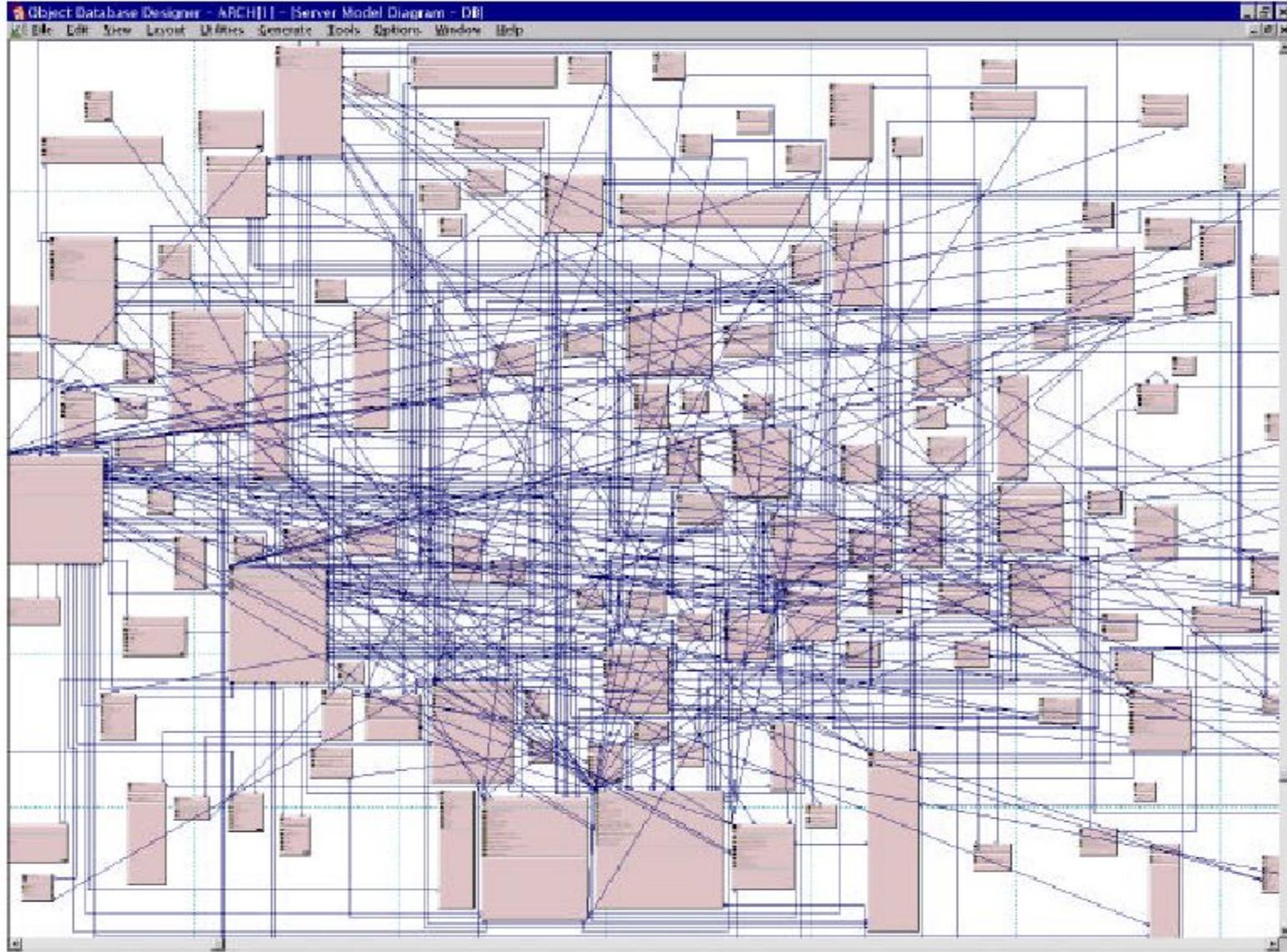
The World Wide Web Consortium (W3C)

Рекомендации 2006 – 2007 годов.

Работа со сложными структурами в РСУБД

- Разложение структур на отношения и сборка их
- Аналог.
 - Автомобиль в гараж – разборка вплоть до гаек и винтиков
 - Из гаража – сборка

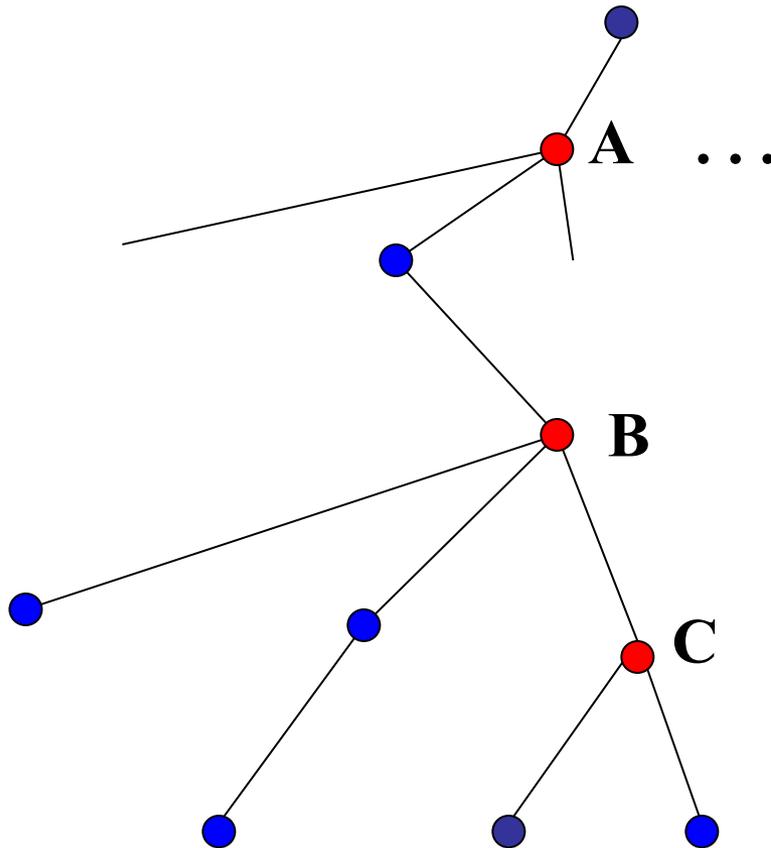
Реляционная модель



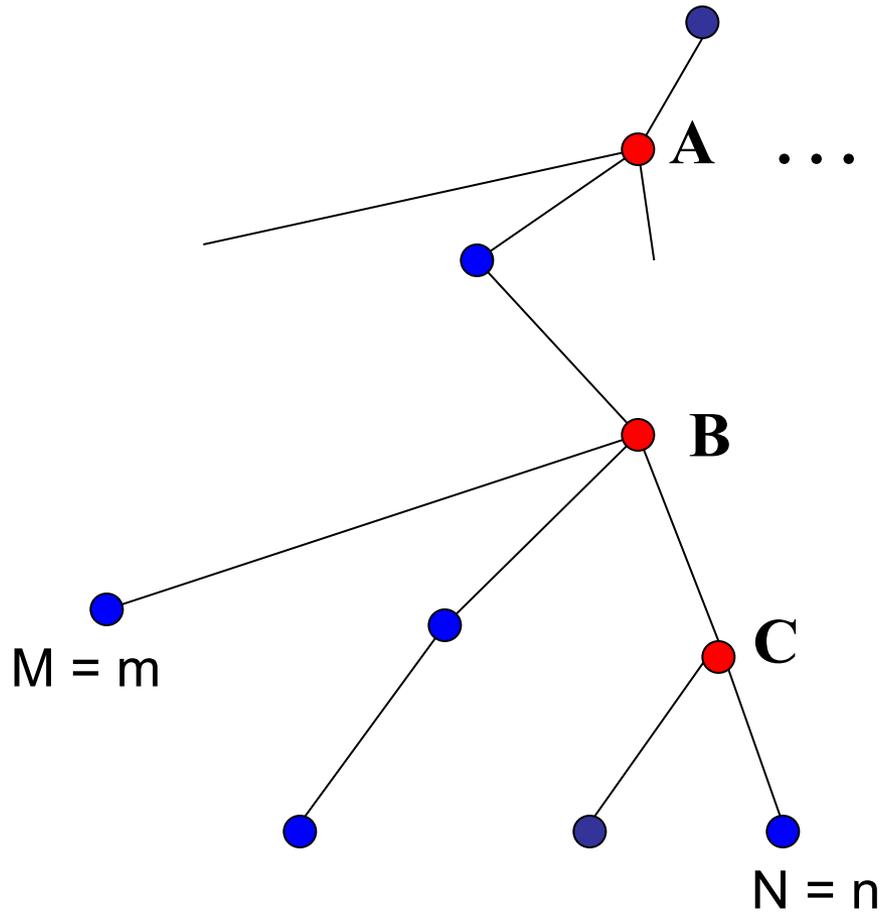
XQuery \supset XPath

- Эта часть языка позволяет описать сложные структуры

a) Отобразить объекты A, которые содержат путь B, C

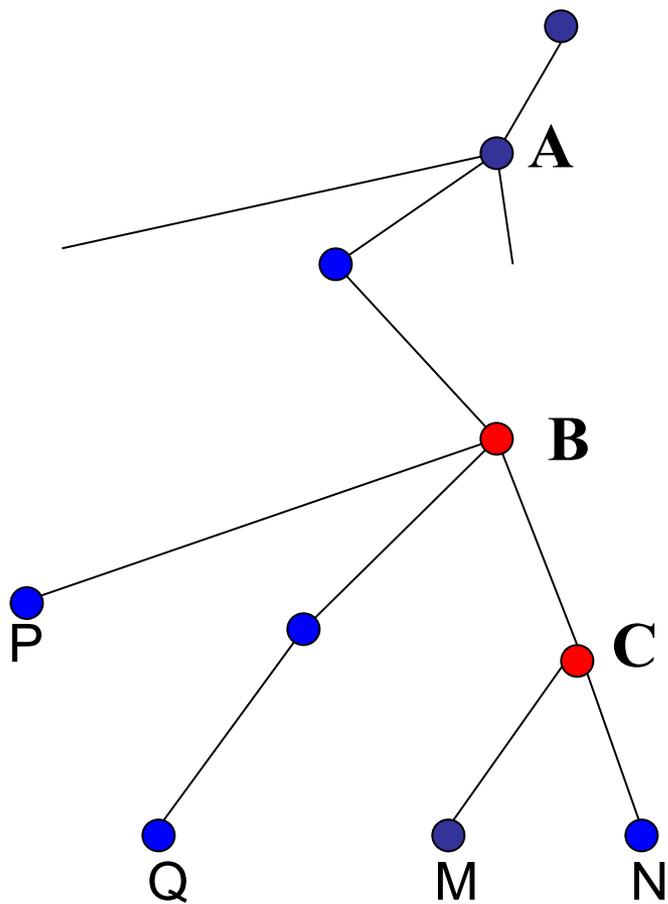


б) Отобрать объекты A , в которых содержат $M = m$, $N = n$

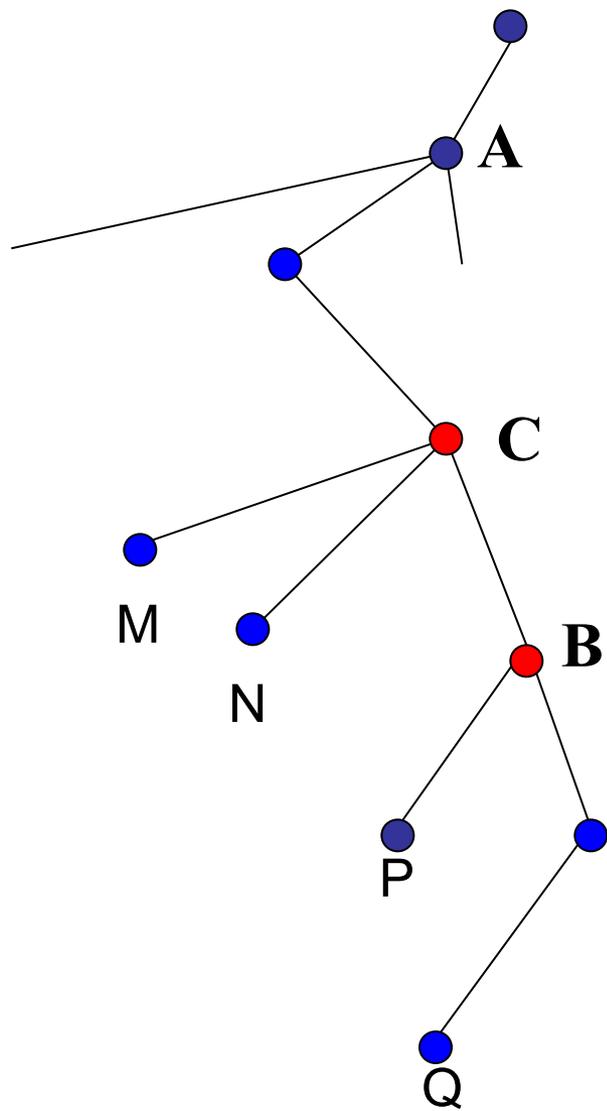


7.4.3. Отображение структуры объекта, реструктуризация

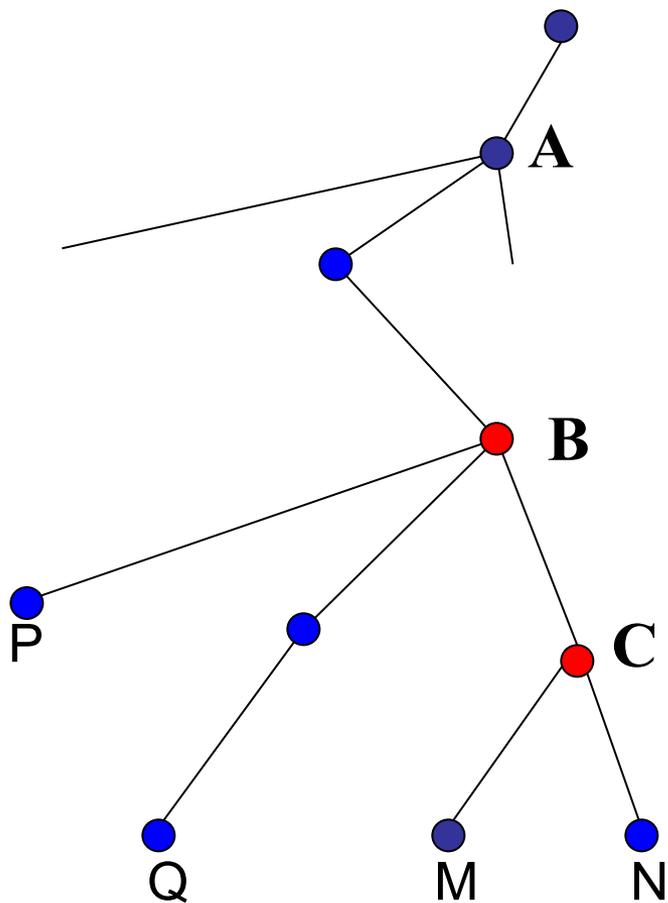
- XSLT (XML Style Language Transformations). W3C
- Язык изменения стиля XML



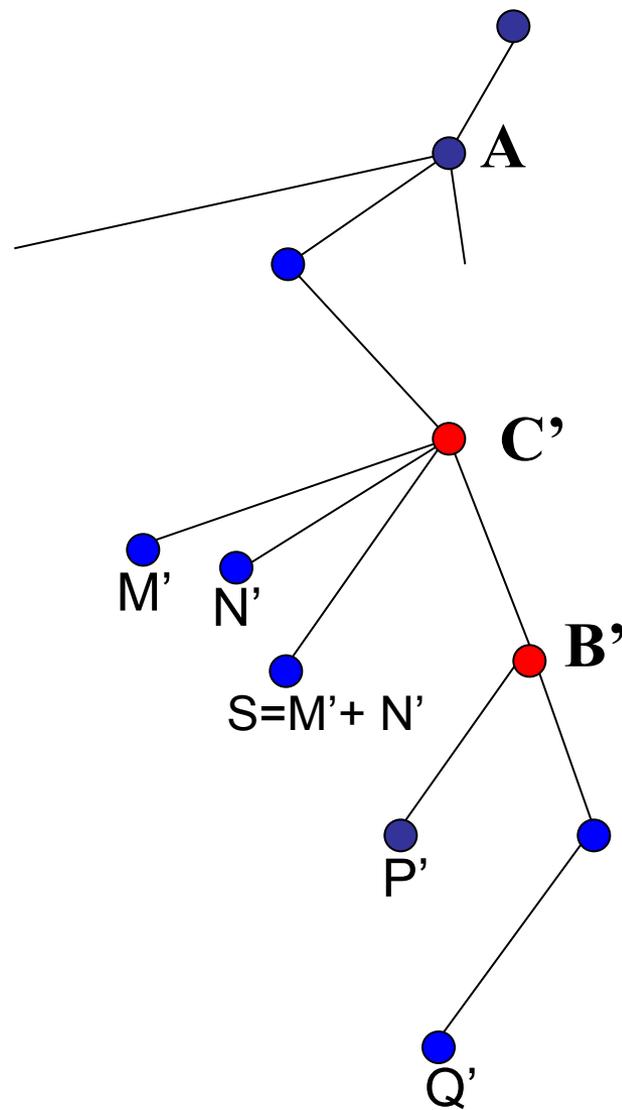
XSLT



Изменение имен, вычисления и др.



XSLT

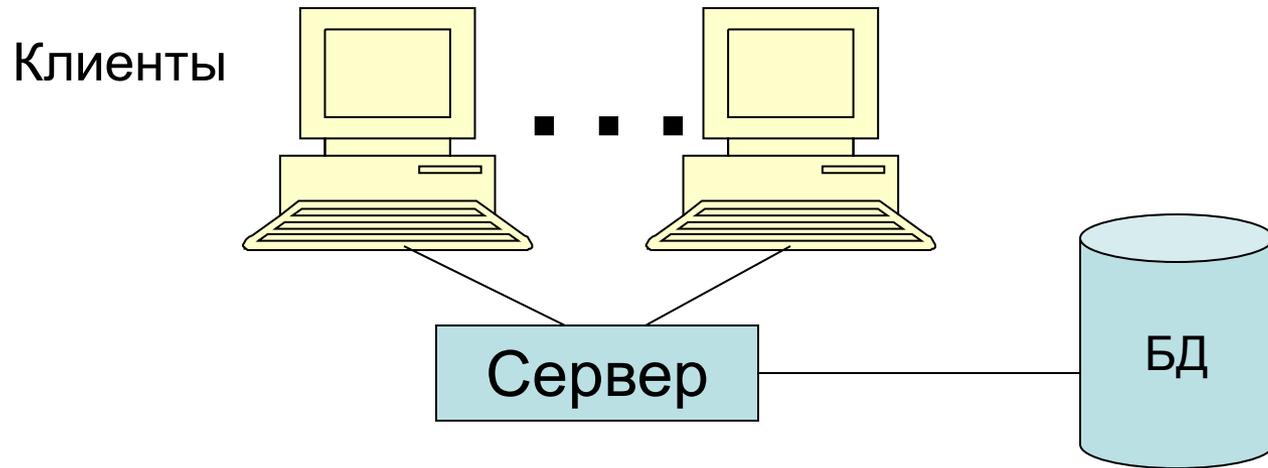


8. Технология эксплуатации

8.1. Потребности пользователей

- 1) Неоднородность (техники, информации - те же объекты названы по разному и др.)
- 2) Распределенность (общие данные разбиты на части)
- 3) Интеграция (Data Warehouse – единые хранилище, индекс)
- 4) Перепроектирование (Reengineering)
- 5) Миграция / наследование (Legacy)
- 6) Интероперабельность (Interoperable) – совместное использование

8.2. Работа в сетях

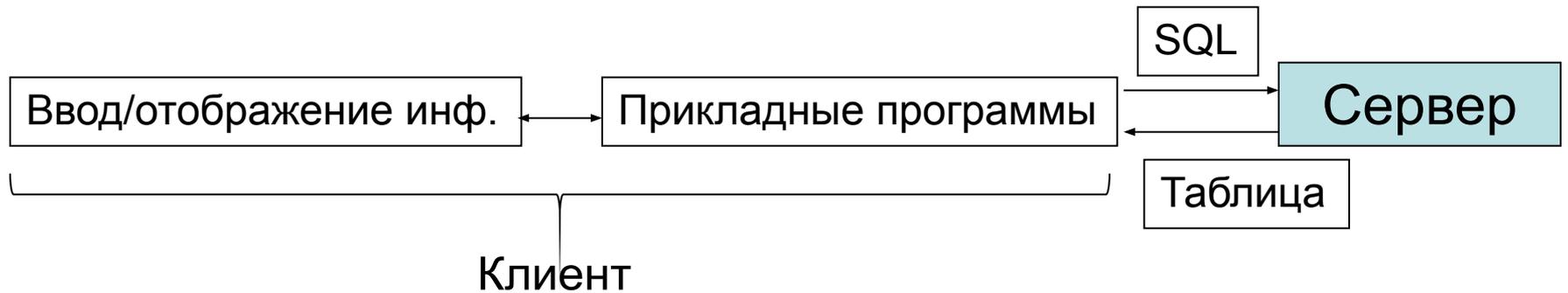


Клиент – серверная модель:

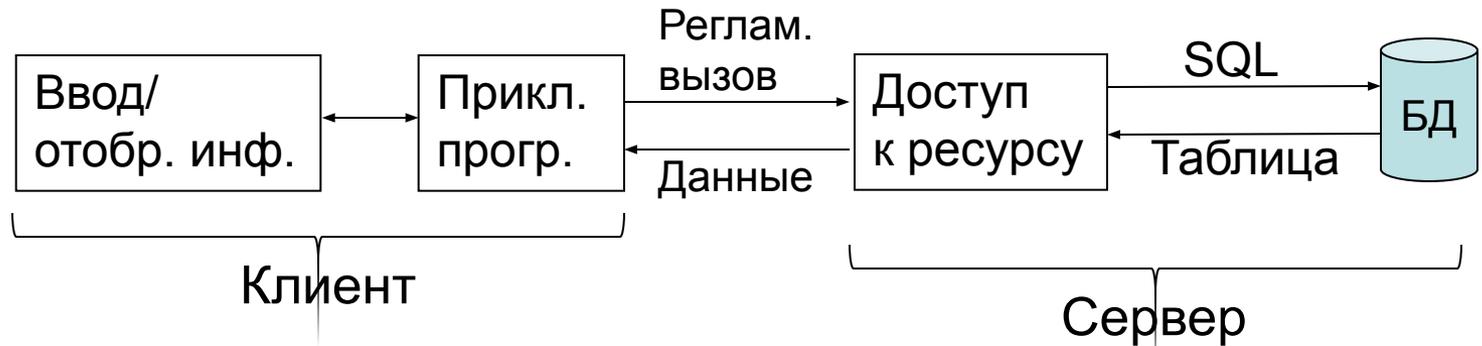
1) Файловый сервер FS



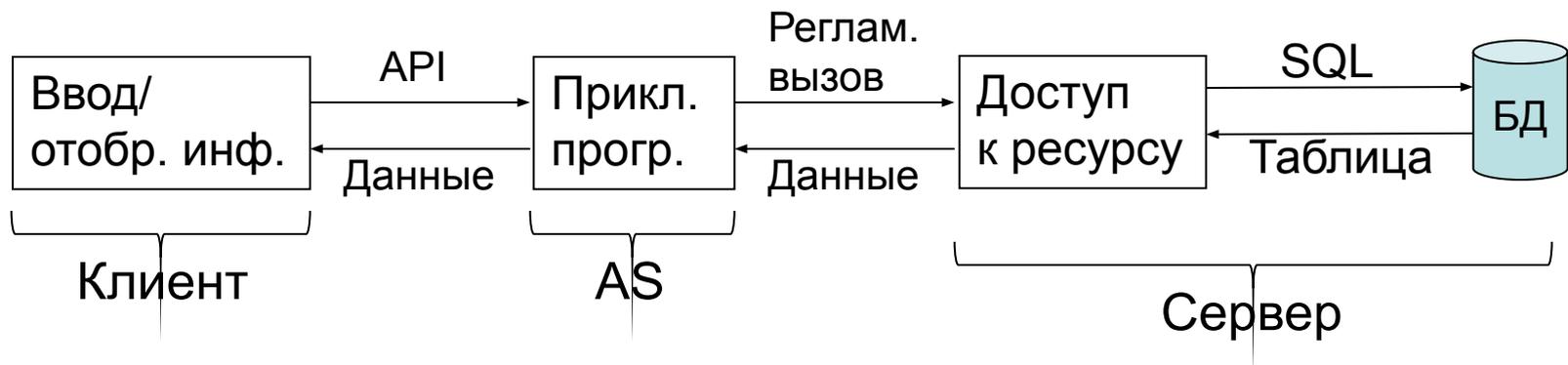
2) Доступ к удаленным ресурсам (RDA – Remove Data Access)



3) Сервер БД (DBS)



4) Сервер приложений (AS – Application Server)



- Системы 1 – 3 двухуровневые
- Система 4 - трехуровневая
- Достоинство 3-уровневой – возможность работы с большим числом клиентов (более 1000)
- Толстый клиент в системах 1 и 2
- Тонкий клиент в системах 4 и 3
- Идеальный тонкий клиент стандартный браузер (MS Explorer)
- Замечательное свойство тонкого клиента – его не нужно устанавливать при изменении программ

8.3. Восстановление после сбоев

8.3.1. Основные понятия

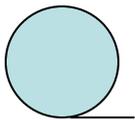
- 1) *Транзакция* – непрерывная деятельность, которая переводит БД из одного согласованного состояния в другое. Транзакция выполняется целиком или вообще не выполняется, т.к. ее нельзя прерывать. В случае сбоя нужно откатиться на начало.
- 2) *Мягкий сбой* (диск – носитель БД – цел). *Жесткий сбой* – диск испорчен.
- 3) *Контрольная точка* – выделенный момент, когда БД в согласованном состоянии.
- 4) *Архив БД* – копия БД, соответствующая некоторой контрольной точке

8.3.2. Ведение архива БД

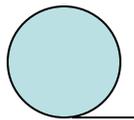
При работе с БД необходимо вести архив, т.к. возможны жесткий сбой, пожар и т.д.

Обычное ведение архива «в кружок»: сын, отец, дед.

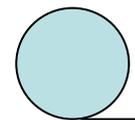
Если время снятия копии БД неделя.



Современное
состояние - **сын**

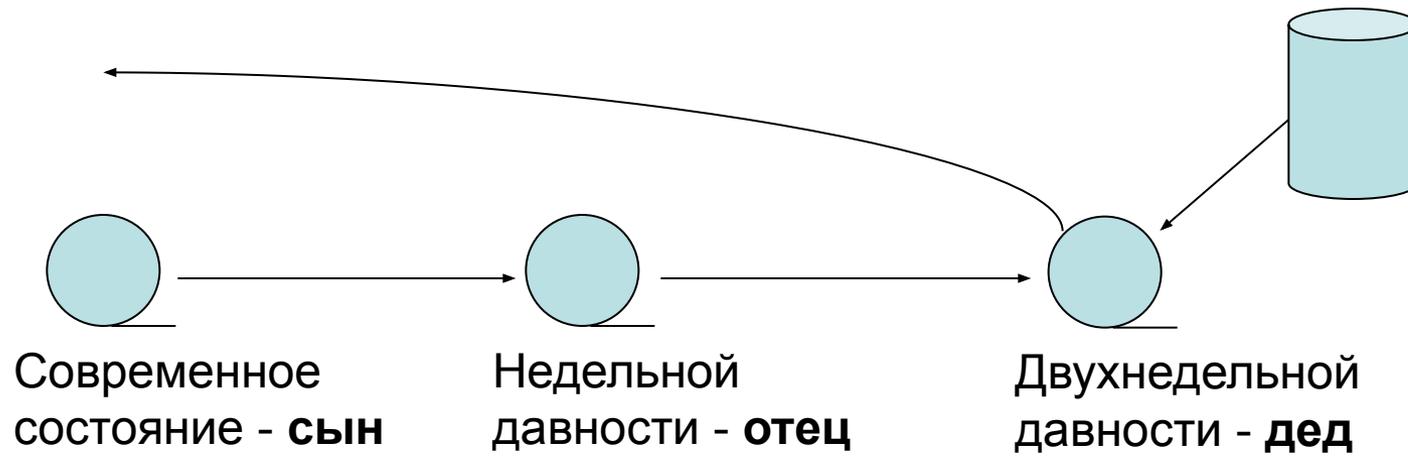


Недельной
давности - **отец**

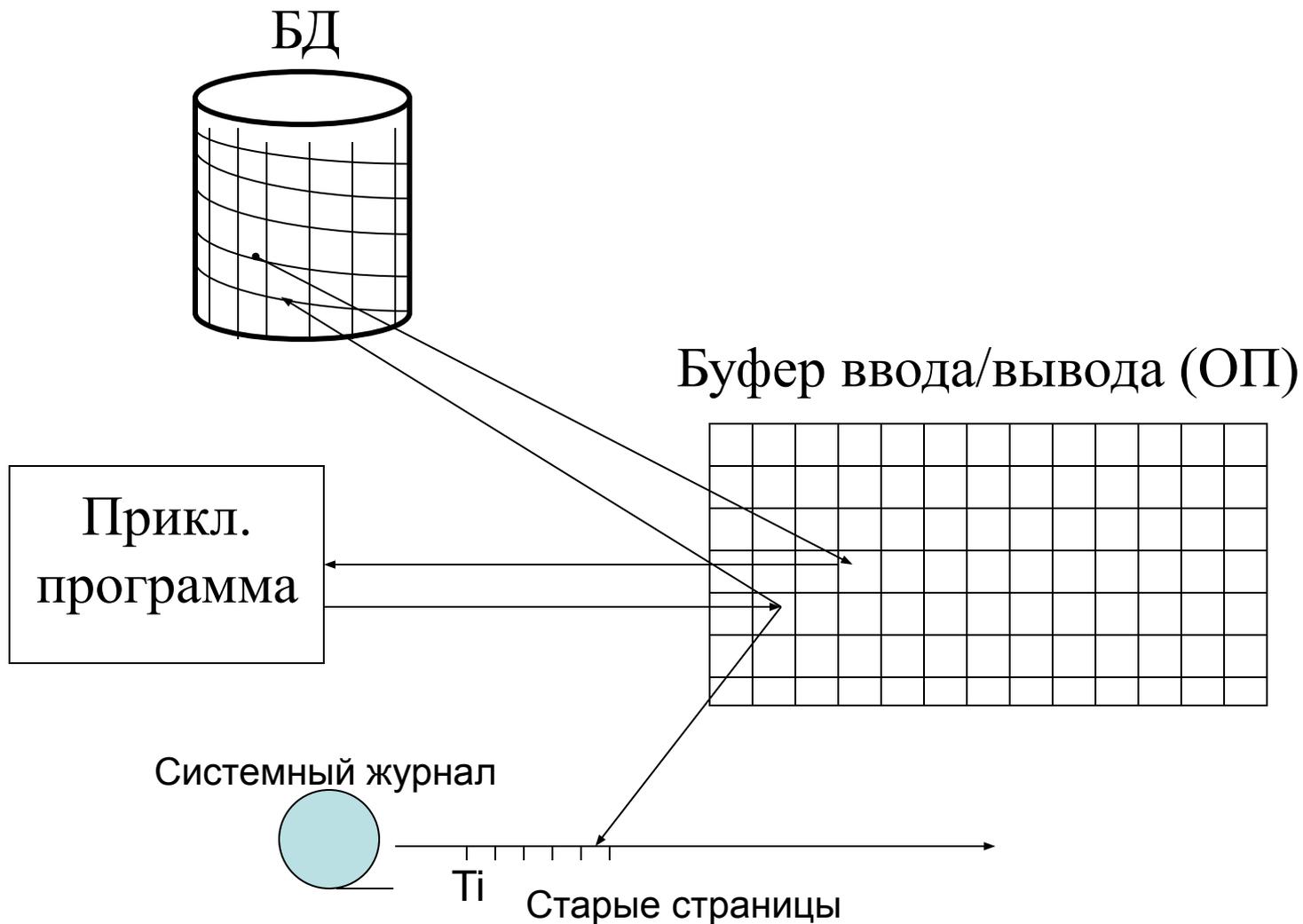


Двухнедельной
давности - **дед**

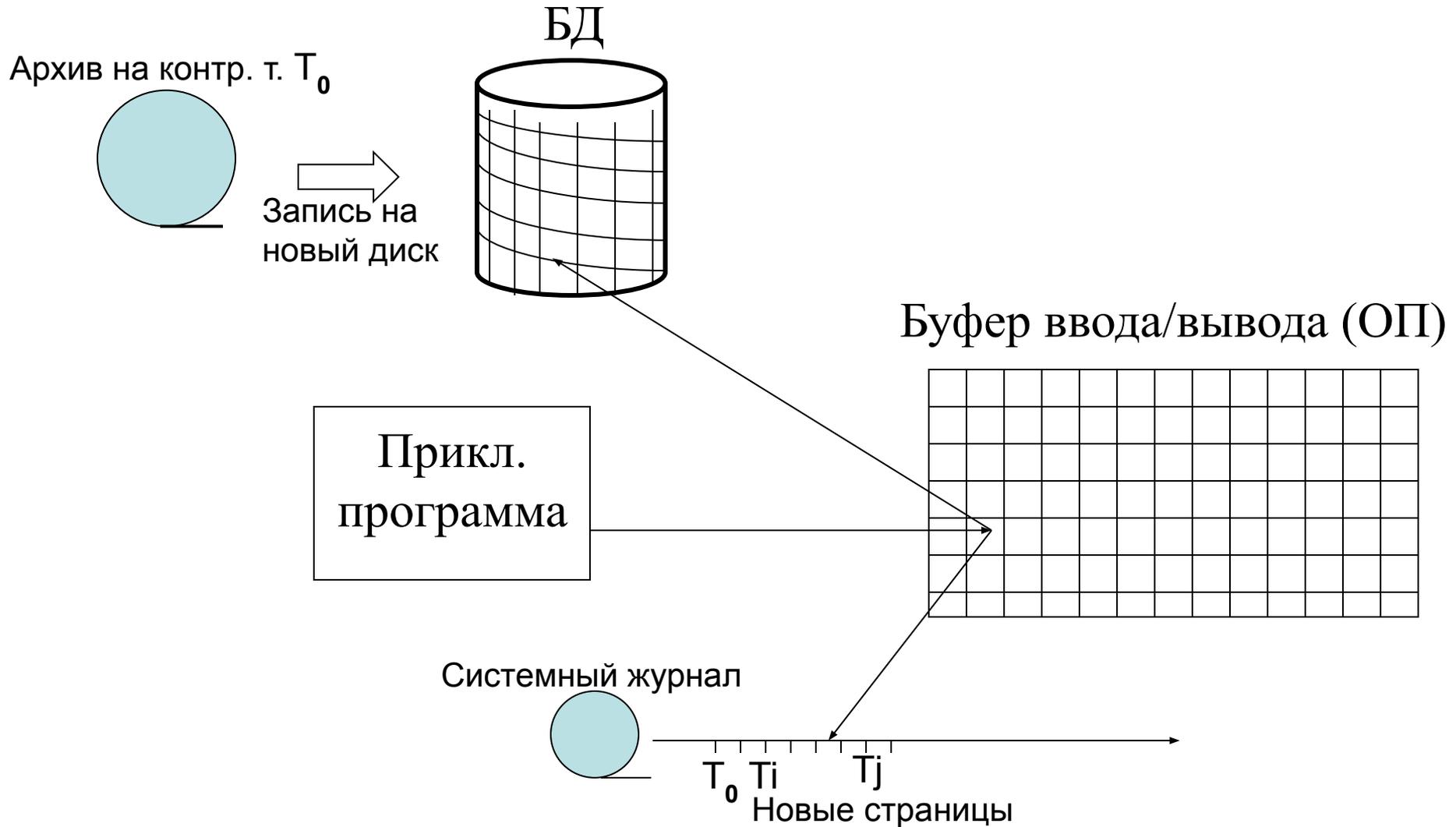
Через неделю при снятии копии



8.3.3. Восстановление назад. Откат системы для мягких сбоев.



8.3.4. Восстановление вперед. Для жестких сбоев.



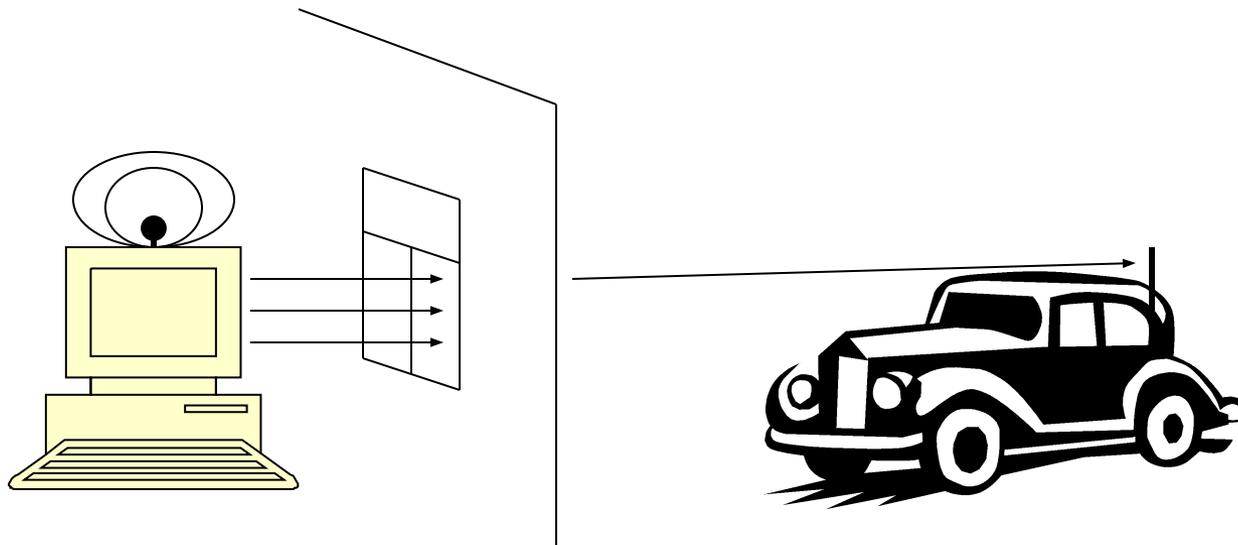
8.4. Защита информации от НСД (несанкционированного доступа)

8.4.1. Основные понятия

- 1) *Пароль* – до 12 символов (помнит человек)
- 2) *Уровни секретности* (от секретно, до Гос. тайны)
 - *Правила назначения пароля:*
 - *Длина (от 8 до 10 символов)*
 - *Ск. букв и ск. цифр*
 - *Запрещенные слова*
 - *Частота замены*
- 3) *Ключи* – от 256 до 4 Кб (человек не помнит). Ключи выдает ФСБ.

4) Защита от излучения

- железные экраны на окнах
- «шумелки»



8.4.2. Закрытые и открытые ключи

Пользователь	Открытый ключ	Закрытый ключ
Иванов	Кл. отк. 1	Кл. закр. 1
Петров	Кл. отк. 2	Кл. закр. 2
.....

Доступны всем
по login
и паролю

Имеет
только
пользователь

8.4.3. Построение открытых и закрытых ключей

Алгоритм RSA

p и q – большие простые числа

$$n = pq$$

Для каждого $e < n$ и взаимно простого с $\phi = (p-1)(q-1)$ существует единственное d такое, что $de \equiv 1 \pmod{\phi}$

(n, e) – открытый ключ

(n, d) – закрытый ключ

$$T^e \pmod{n} = \underline{T} \quad T < n$$

$$\underline{T}^d \pmod{n} = T$$

Стойкость алгоритма – ск. операций нужно произвести, чтобы расшифровать.

Стойкость алгоритм RSA определяется трудностью разложения на простые множители числа $n = pq$

Сколько существует простых чисел?

Простых чисел бесконечно много.

Док-во Евклида (3-й век до н.э.)

Допустим, что количество простых чисел конечно. Перемножим их и прибавим единицу. Полученное число не делится ни на одно из конечного набора простых чисел (остаток единица). Значит, оно новое простое или должно делиться на некоторое простое число, не включённое в этот набор.

**Наибольшее известное простое.
Сколько десятичных знаков ?**

9.808.358 десятичных знаков -
10 томов по 1000 страниц.
44-е известное простое

число Мерсенна = $2^{32582657} - 1$.

(найдено в 2006 г.)

В 2005 найдено из 9.152.052 цифр.

Фирма EFF платит
\$100.000 за каждое новое простое
число из более чем
 10^7 десятичных цифр.

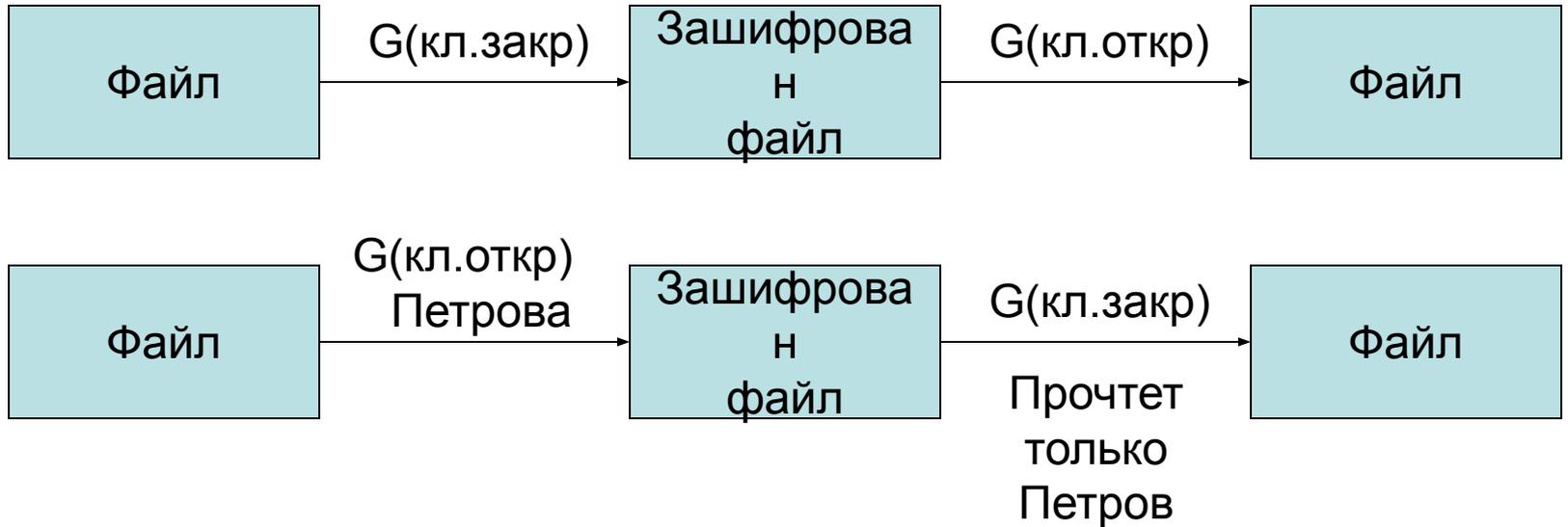
8.4.4. Электронная подпись



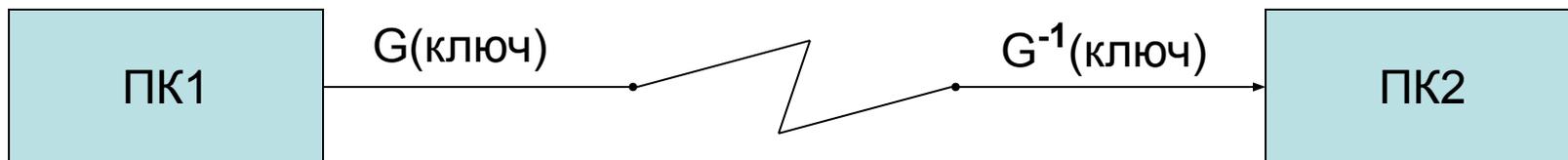
Текст не изменяется от подписи

8.4.5. Шифрование файлов, каналов и устройств

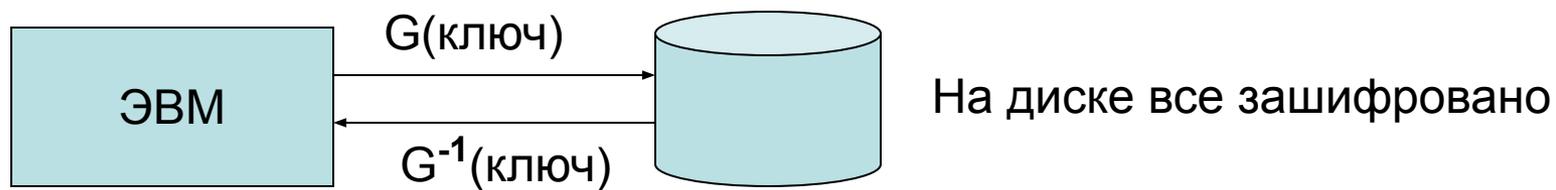
1) Файлы



2) Шифрование каналов

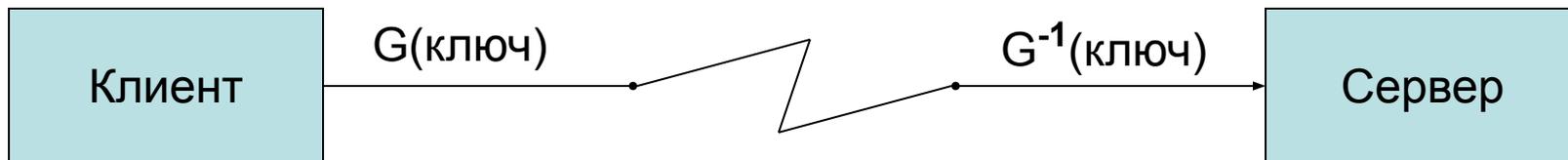


3) Шифрование устройств



4) Шифрование сеанса связи MS

(login + пароль) – начало сеанса



Ключ – случайная величина **К** из счетчик битов

$\mathbf{K}^e = \underline{\mathbf{K}}$ $\underline{\mathbf{K}}$ посылается с сервера клиенту

$\underline{\mathbf{K}}^d = \mathbf{K}$ расшифровывается на клиенте

Шифровка ведется по кускам (гумирование) при помощи ключа **К**

Текст



гумирование