



# БАЗЫ ДАННЫХ И ЭКСПЕРТНЫЕ СИСТЕМЫ

ФИО преподавателя: Тараканов О.В., канд. техн. наук, доцент  
e-mail: [tarakanov@mirea.ru](mailto:tarakanov@mirea.ru)

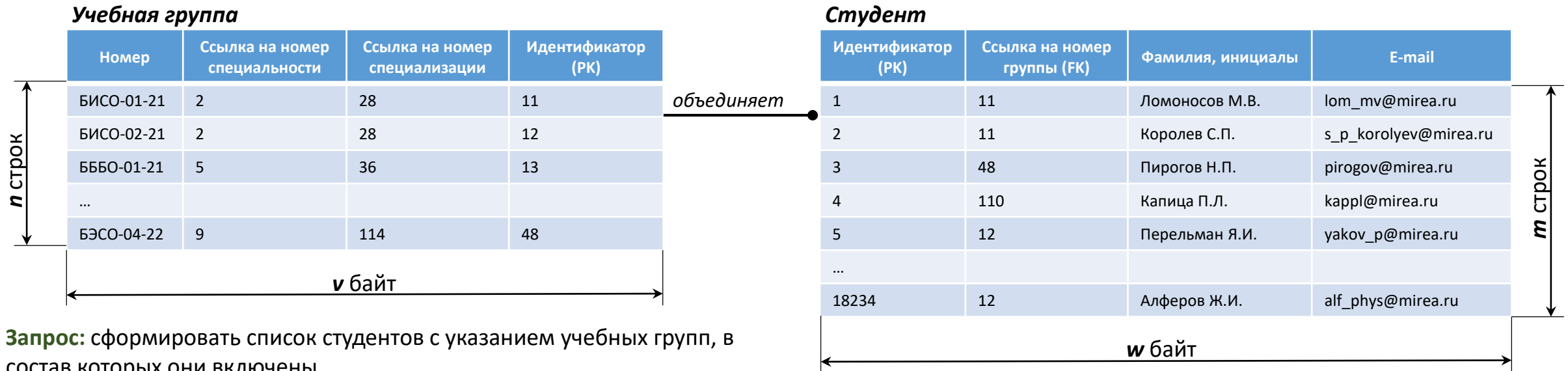
# ОПТИМИЗАЦИЯ БАЗЫ ДАННЫХ

Учебные вопросы:

1. Функциональный тюнинг
2. Структурный тюнинг
3. Организация технической инфраструктуры

# Функциональный тюнинг

## Понятие тюнинга базы данных



**Запрос:** сформировать список студентов с указанием учебных групп, в состав которых они включены.

NESTED LOOP

### Методы повышения скорости выборки:

1. *Повышение производительности вычислительной платформы* (высокоскоростные (дорогие) комплектующие сервера – центральный процессор, ОЗУ и внешняя память).
2. *Организация вложения цикла с большим числом итераций* – меньшее число открытий файла данных (медленные операции).
3. *Определение рационального размера файла данных* – исключение страничности, уменьшение объемов своппирования.
4. *Организация рационального исчисления предписаний запроса* – уменьшение числа «медленных» операций, определение наиболее выигрышной последовательности действий.
5. *Реструктуризация данных* – денормализация, сортировка, индексирование, организация выгрузки в память сервера, группирование файлов данных по назначению, управление блоками базы данных, расширениями...

## Рациональная структура базы данных

**Учебная группа**

Номер	Ссылка на номер специальности	Ссылка на номер специализации	Идентификатор (РК)
БИСО-01-21	2	28	11
БИСО-02-21	2	28	12
БББО-01-21	5	36	13
...			
БЭСО-04-22	9	114	48

объединяет

**Студент**

Идентификатор (PK)	Ссылка на номер группы (FK)	Фамилия, инициалы	E-mail
1	11	Ломоносов М.В.	lom_mv@mirea.ru
2	11	Королев С.П.	s_p_korolyev@mirea.ru
3	48	Пирогов Н.П.	pirogov@mirea.ru
4	110	Капица П.Л.	kappl@mirea.ru
5	12	Перельман Я.И.	yakov_p@mirea.ru
...			
18234	12	Алферов Ж.И.	alf_phys@mirea.ru

## Денормализация – исключение вложенных циклов

## Индексирование

## Предварительное соединение таблиц

**Студент\_ДНФ**

Идентификатор (РК)	Номер группы	Фамилия, инициалы	E-mail	Ссылка на номер специальности	Ссылка на номер специализации
1	БИСО-01-21	Ломоносов М.В.	lom_mv@mirea.ru	2	28
2	БИСО-01-21	Королев С.П.	s_p_korolyev@mirea.ru	2	28
3	БЭСО-04-22	Пирогов Н.П.	pirogov@mirea.ru	9	114
4	БАСО-02-20	Капица П.Л.	kappl@mirea.ru	11	96
5	БИСО-02-21	Перельман Я.И.	yakov_p@mirea.ru	2	28
...					
18234	БИСО-02-21	Алферов Ж.И.	alf_phys@mirea.ru	2	28

**Индекс (Студент)**

Фамилия, инициалы	Адрес блока данных
Ломоносов М.В.	h178A20FFD0200
Королев С.П.	h056AAF3C240D1
Пирогов Н.П.	h1690DAA2AB944
Капица П.Л.	h102A45CC23BAF
Перельман Я.И.	h11FF230AD56C
	↓
Алферов Ж.И.	h19923FAB69CC01

Файл  
данных

**Представление (Материализованное представление)**

[illegible]

# Функциональный тюнинг

## План выполнения запроса, EXPLAIN PLAN

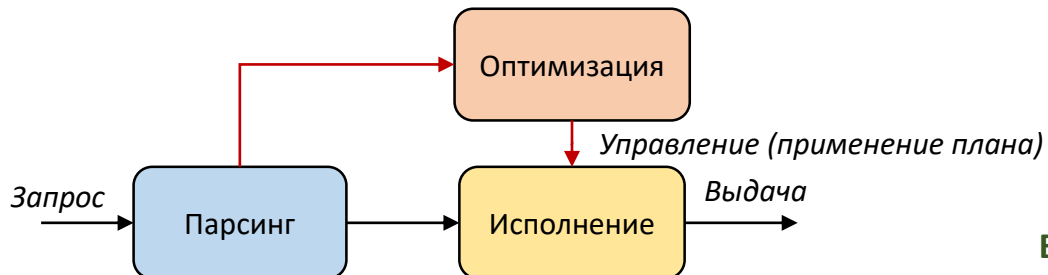
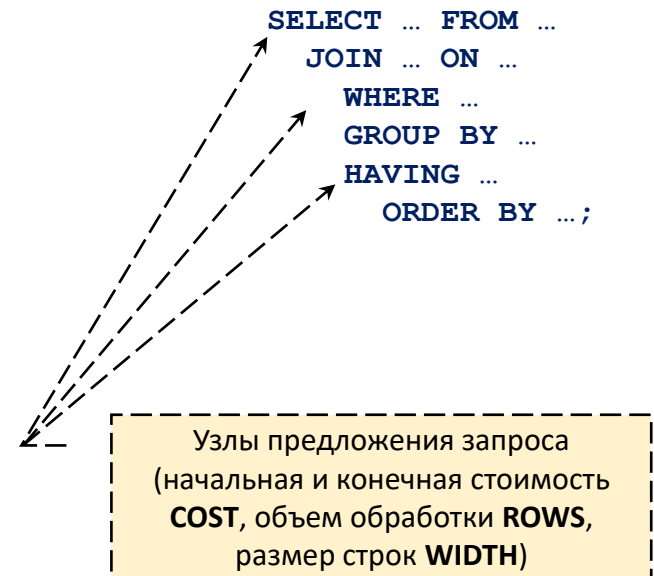
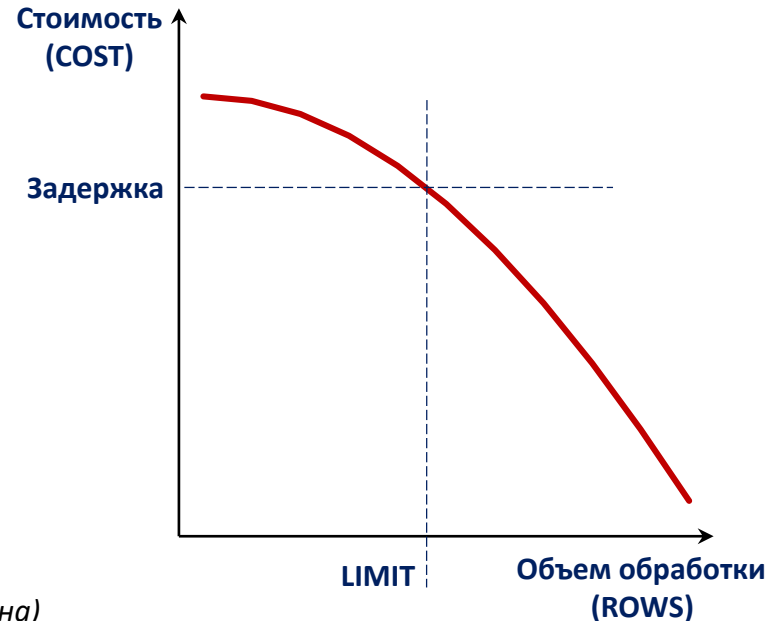


### Параметры запроса:

**Стоимость** (англ. – COST) – это относительная величина задержки до момента, когда начнется вывод данных выборки, и момента, когда будут выведены все отобранные строки.

**Объем обработки** (англ. – ROWS) – это ожидаемое число строк, которые будут отобраны при полном выполнении текущего этапа запроса.

**Размер строк** (англ. – WIDTH) – это ожидаемый средний размер в байтах выводимых строк.



### Вывод параметров плана выполнения запроса:

**EXPLAIN <SQL-предложение запроса>;**

# Функциональный тюнинг

Оптимизация базы данных сортировкой

Задача: Соединить явно заданные таблицы (вложенный цикл)

Team

ID	Команда	Город	...
1	Динамо	Москва	
2	Спартак	Воронеж	
3	Локомотив	Екатеринбург	
4	ЦСКА	Москва	
5	Зенит	Санкт-Петербург	
6	Арсенал	Тула	

NESTED LOOP

Player

ID	Игрок	Квалификация	...
5	Краснов В.А.	МС России	
2	Онопко В.Г.	МСМК	
2	Беланов С.И.	ЗМС	
5	Мунтян А.С.	ЗМС	
1	Блохин О.Н.	МСМК	
3	Баль О.А.	МС России	

Для каждой строки одной из таблиц (ведущей) путем тотального перебора выполняется поиск в другой таблице (ведомой) строк, соответствующих условию соединения. Операция повторяется для всех строк ведущей таблицы.

Team join Player

Team.ID	Player.ID	Команда	Город	Игрок	Квалификация	...
1	1	Динамо	Москва	Блохин О.Н.	МСМК	
2	2	Спартак	Воронеж	Онопко В.Г.	МСМК	
2	2	Спартак	Воронеж	Беланов С.И.	ЗМС	
3	3	Локомотив	Екатеринбург	Баль О.А.	МС России	
5	5	Зенит	Санкт-Петербург	Краснов В.А.	МС России	
5	5	Зенит	Санкт-Петербург	Мунтян А.С.	ЗМС	

# Функциональный тюнинг

Задача: Соединить явно заданные таблицы (вложенный цикл с индексированием – сортировка)

Team

ID	Команда	Город	...
1	Динамо	Москва	
2	Спартак	Воронеж	
3	Локомотив	Екатеринбург	
4	ЦСКА	Москва	
5	Зенит	Санкт-Петербург	
6	Арсенал	Тула	



Player

ID	Игрок	Квалификация	...
5	Краснов В.А.	МС России	
2	Онопко В.Г.	МСМК	
2	Беланов С.И.	ЗМС	
5	Мунтян А.С.	ЗМС	
1	Блохин О.Н.	МСМК	
3	Баль О.А.	МС России	

Для каждой строки одной из таблиц (ведущей) путем тотального перебора выполняется поиск в другой таблице (ведомой) строк, соответствующих условию соединения. Ведомая таблица предварительно отсортирована, например, по убыванию значений. Операция повторяется для всех строк ведущей таблицы.

Team join Player						
Team.ID	Player.ID	Команда	Город	Игрок	Квалификация	...
1	1	Динамо	Москва	Блохин О.Н.	МСМК	
2	2	Спартак	Воронеж	Онопко В.Г.	МСМК	
2	2	Спартак	Воронеж	Беланов С.И.	ЗМС	
3	3	Локомотив	Екатеринбург	Баль О.А.	МС России	
5	5	Зенит	Санкт-Петербург	Краснов В.А.	МС России	
5	5	Зенит	Санкт-Петербург	Мунтян А.С.	ЗМС	

# Функциональный тюнинг

Задача: Соединить явно заданные таблицы (слияние списков)

Team

ID	Команда	Город	...
1	Динамо	Москва	
2	Спартак	Воронеж	
3	Локомотив	Екатеринбург	
4	ЦСКА	Москва	
5	Зенит	Санкт-Петербург	
6	Арсенал	Тула	

Player

ID	Игрок	Квалификация	...
5	Краснов В.А.	МС России	
2	Онопко В.Г.	МСМК	
2	Беланов С.И.	ЗМС	
5	Мунтян А.С.	ЗМС	
1	Блохин О.Н.	МСМК	
3	Баль О.А.	МС России	



Входные таблицы должны быть отсортированы по столбцам, участвующим в условии соединения. Соединение осуществляется за одно сканирование (проход по) каждой из входных таблиц.

Team join Player						
Team.ID	Player.ID	Команда	Город	Игрок	Квалификация	...
1	1	Динамо	Москва	Блохин О.Н.	МСМК	
2	2	Спартак	Воронеж	Онопко В.Г.	МСМК	
2	2	Спартак	Воронеж	Беланов С.И.	ЗМС	
3	3	Локомотив	Екатеринбург	Баль О.А.	МС России	
5	5	Зенит	Санкт-Петербург	Краснов В.А.	МС России	
5	5	Зенит	Санкт-Петербург	Мунтян А.С.	ЗМС	



# Функциональный тюнинг

Задача: Соединить явно заданные таблицы (хеширование)

Team

ID	Команда	Город	...
1	Динамо	Москва	
2	Спартак	Воронеж	
3	Локомотив	Екатеринбург	
4	ЦСКА	Москва	
5	Зенит	Санкт-Петербург	
6	Арсенал	Тула	

Player

ID	Игрок	Квалификация	...
5	Краснов В.А.	МС России	
2	Онопко В.Г.	МСМК	
2	Беланов С.И.	ЗМС	
5	Мунтян А.С.	ЗМС	
1	Блохин О.Н.	МСМК	
3	Баль О.А.	МС России	

Меньшая из соединяемых таблиц помещается в структуру данных в памяти – хеш-таблицу, которая обеспечивает очень высокую скорость поиска. Затем для каждой строки из большей таблицы выполняется поиск значений, соответствующих условию соединения. Результаты помещаются в выходную таблицу.

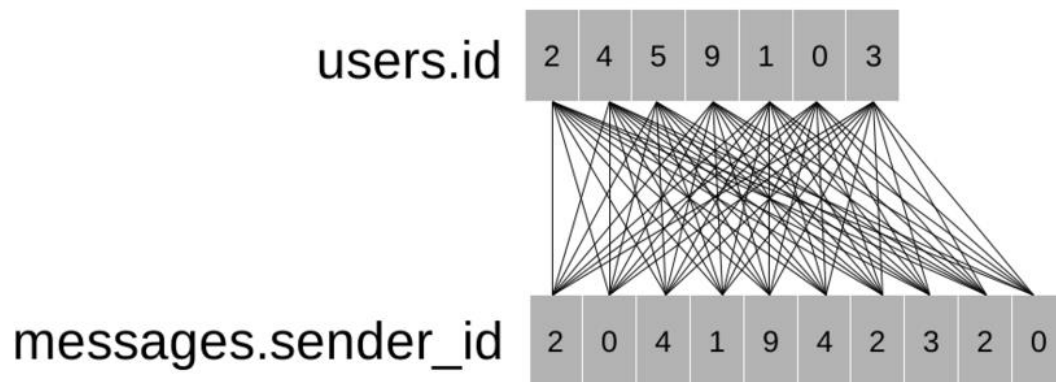


Team join Player

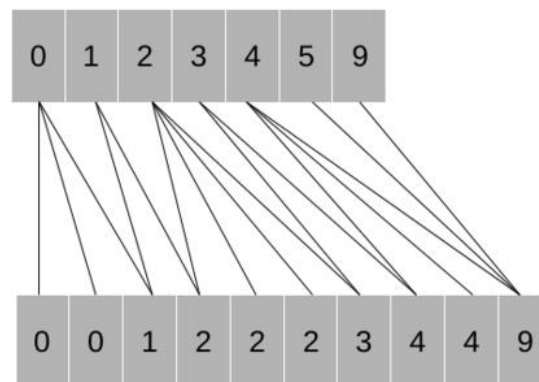
Team.ID	Player.ID	Команда	Город	Игрок	Квалификация	...
1	1	Динамо	Москва	Блохин О.Н.	МСМК	
2	2	Спартак	Воронеж	Онопко В.Г.	МСМК	
2	2	Спартак	Воронеж	Беланов С.И.	ЗМС	
3	3	Локомотив	Екатеринбург	Баль О.А.	МС России	
5	5	Зенит	Санкт-Петербург	Краснов В.А.	МС России	
5	5	Зенит	Санкт-Петербург	Мунтян А.С.	ЗМС	

# Функциональный тюнинг

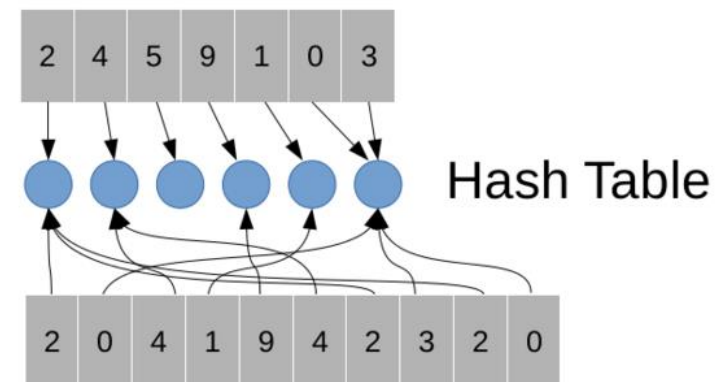
Nested Loop Join



Merge Join



Hash Join



Скорость соединения	низкая	высокая	очень высокая
Затраты памяти	незначительные	низкие	очень высокие
Накладные расходы	отсутствуют	предварительная сортировка соединяемых таблиц	определение меньшей по числу строк таблицы, построение хеш-таблицы, разрешение коллизий

# Функциональный тюнинг

## Оптимизация базы данных с помощью индексов

### Индекс (Студент)

Фамилия, инициалы	Адрес блока данных
Ломоносов М.В.	h178A20FFD0200
Королев С.П.	h056AAF3C240D1
Пирогов Н.П.	h1690DAA2AB944
Капица П.Л.	h102A45CC23BAF
Перельман Я.И.	h11FFF230AD56C
Алферов Ж.И.	h19923FAB67CC01



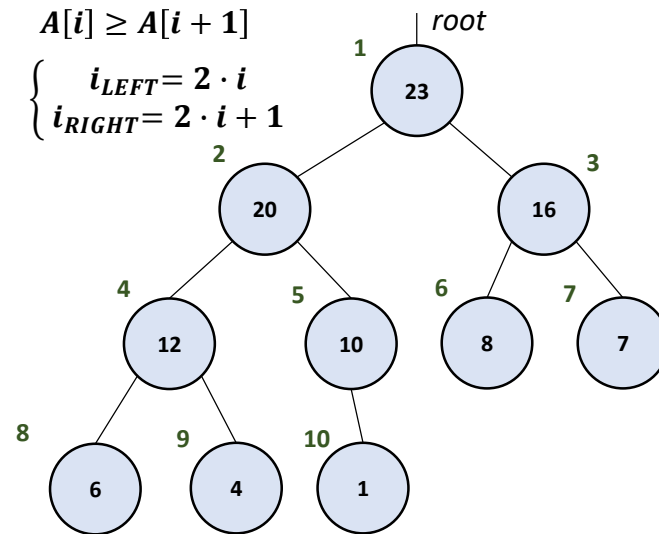
**Индекс** (англ. – INDEX) – специальная конструкция (объект) базы данных, обеспечивающая быстрый поиск данных, особенно, по неключевым атрибутам, когда строки не отсортированы в наиболее выгодном для поиска порядке.

### Список – LIST

	A
1	16
2	23
3	8
4	4
5	1
6	20
7	12
8	7
9	10
10	6

$$T \sim o(n)$$

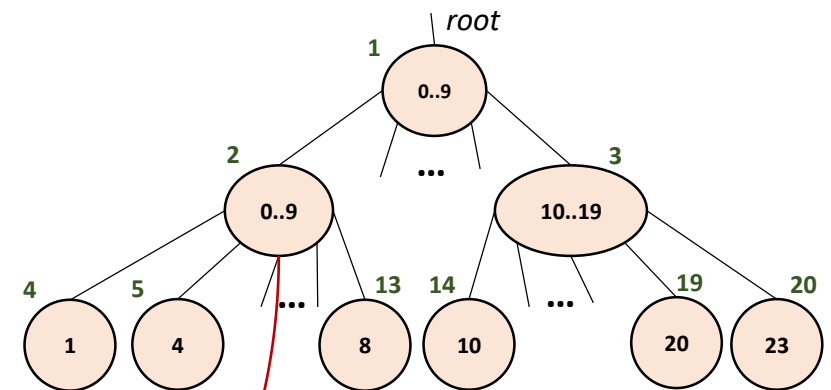
### Двоичная куча – BINARY HEAP



$$T \sim o(\log_2 n)$$

$$T \sim o(\log_t h)$$

### Сбалансированное дерево – B-tree (balanced)



$$t = 10 \quad h = 2$$

$$T \sim o(\log_{10} 2)$$

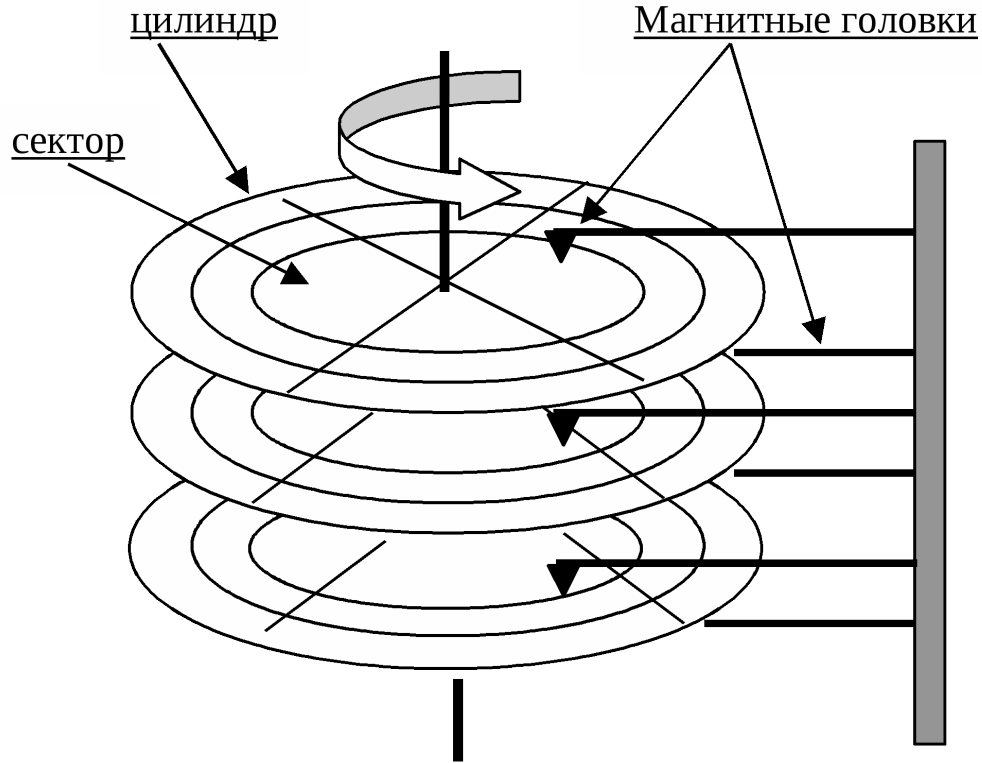
### Хэшированный индекс – HASH INDEX

$$ADDRESS = \varphi(PRIMARY KEY)$$

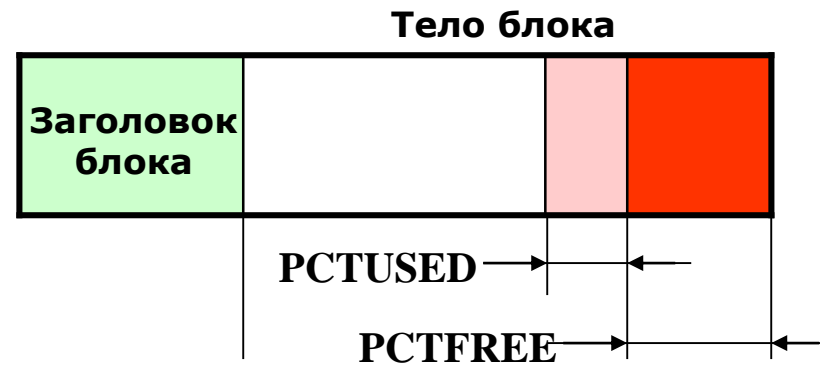
### Битовый индекс – BitMap INDEX



# Структурный тюнинг



**Блок базы данных** (англ. – DBblock) – контейнер, предполагающий одномоментную запись или чтение (соответственно однократное позиционирование головок чтения / записи) содержимого на / из дискового накопителя сервера баз данных.



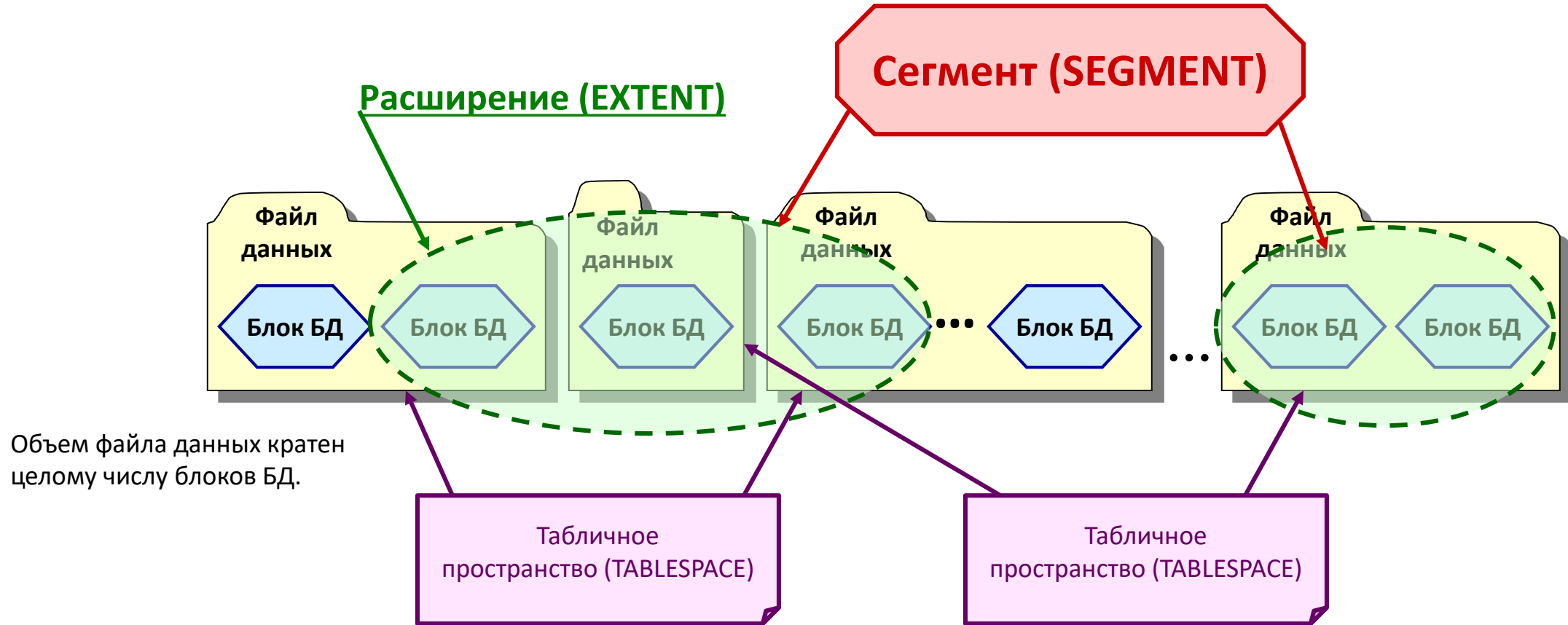
$$V_{\Phi C} \neq V_{ББД}$$

Время чтения 1 страницы текста =  
Среднее время позиционирования головки чтения/записи \*  
( 8 \* Число строк \* Число символов в строке )

Время чтения 1 страницы текста = 9,8 мс \* ( 8 \* 40 \* 80 ) = ?

OLTP-система:  $V_{ББД}$  ↓  
GP-система:  $V_{ББД} = 8...16$  кБ  
OLAP (DSS)-система:  $V_{ББД}$  ↑

# Структурный тюнинг



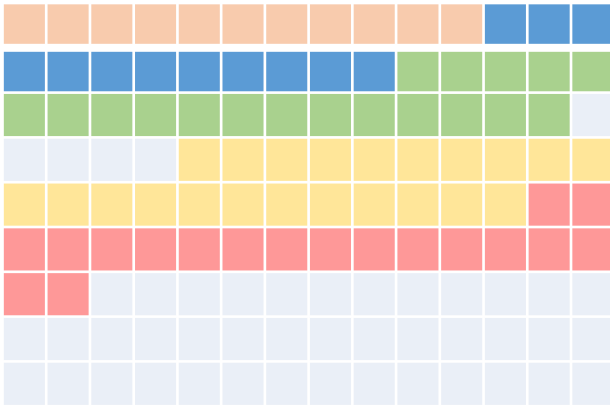
**Расширение** (англ. – EXTENT) – специальная конструкция, логически объединяющая несколько блоков базы данных, для обеспечения их последовательного размещения на поверхности пластины накопителя на магнитных дисках сервера баз данных.

**Сегмент** (англ. – SEGMENT) – специальная конструкция, логически объединяющая несколько расширений, предназначенная для непосредственного хранения пользовательских данных на физическом уровне. Сегменты реализуются в виде таблиц, индексов, представлений, хранимых процедур, функций, триггеров, последовательностей, синонимов и других объектов базы данных.

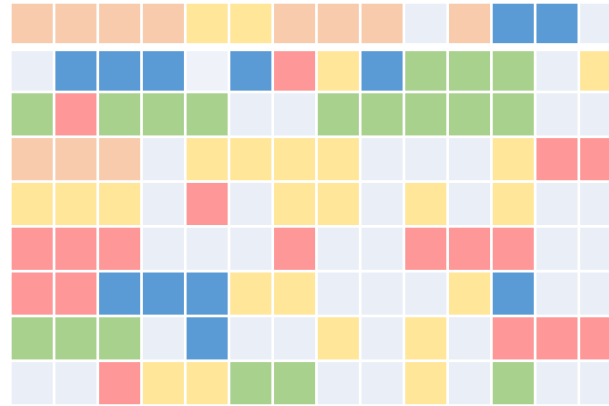
**Табличное пространство** (англ. – TABLESPACE) – специальная конструкция, логически объединяющая несколько файлов данных с одинаковыми методами доступа к ним, обеспечивающая параллельное обращение к байтам разного предназначения (таблицы, индексы, сегменты отката,...).

# Структурный тюнинг

Дефрагментированный диск



Фрагментированный диск



Дефрагментация диска – затратная процедура.

Число позиционирований головки чтения / записи прямо пропорционально числу блоков, в которых содержатся фрагменты файла.

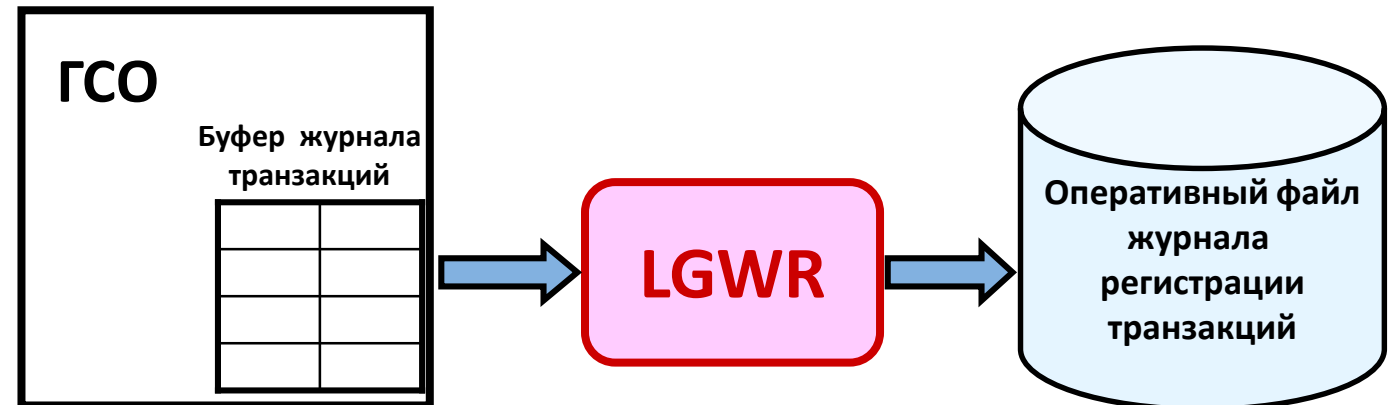
Режим перезаписи байтов обеспечивает неизменную фрагментированность файла, в отличие от режима дозаписи в файл.

$$T_{READ} = N_{BLOCK} \cdot t_1$$

**Инженерный лайфхак (lifehack):** «Чистый» диск сервера баз данных, после форматирования, заполнить «до отказа» файлами данных будущей системы баз данных. Так как файлы данных будут записаны последовательно и поочередно, то фрагментированность диска будет минимальной, а время чтения файлов с него – наименьшим. Не возможность записи на диск новых файлов позволит защитить его от возникновения фрагментированности в процессе эксплуатации.

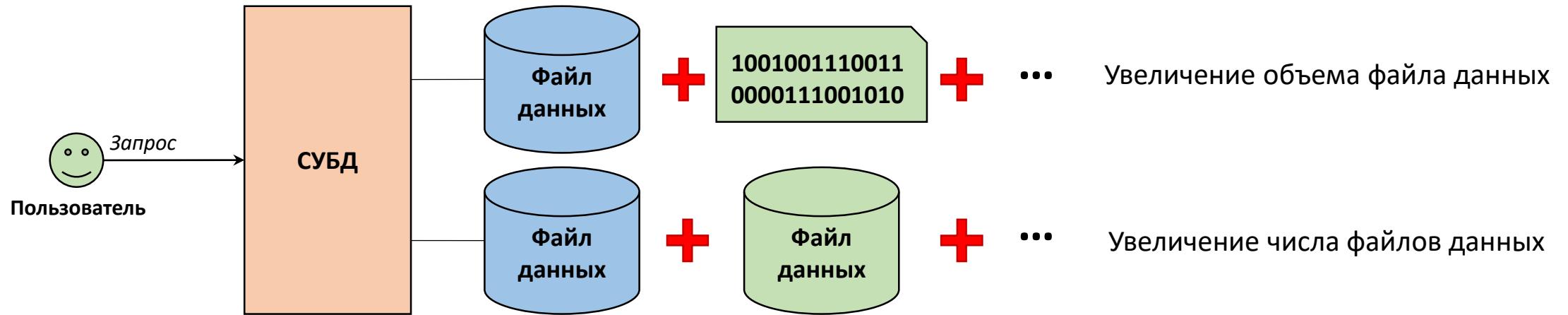
**Журнал регистрации транзакций** – конструкция из файла журнала регистрации транзакций и буфера, хранящего записи в Глобальной системной области (оперативной памяти сервера баз данных), обеспечивающей уменьшение нагрузки на подсистему ввода / вывода служебными данными.

**Инженерный лайфхак (lifehack):** оперативные файлы журнала регистрации транзакций хранить отдельно от файлов данных.



# Организация технической инфраструктуры

## Стратегии управления файлом данных



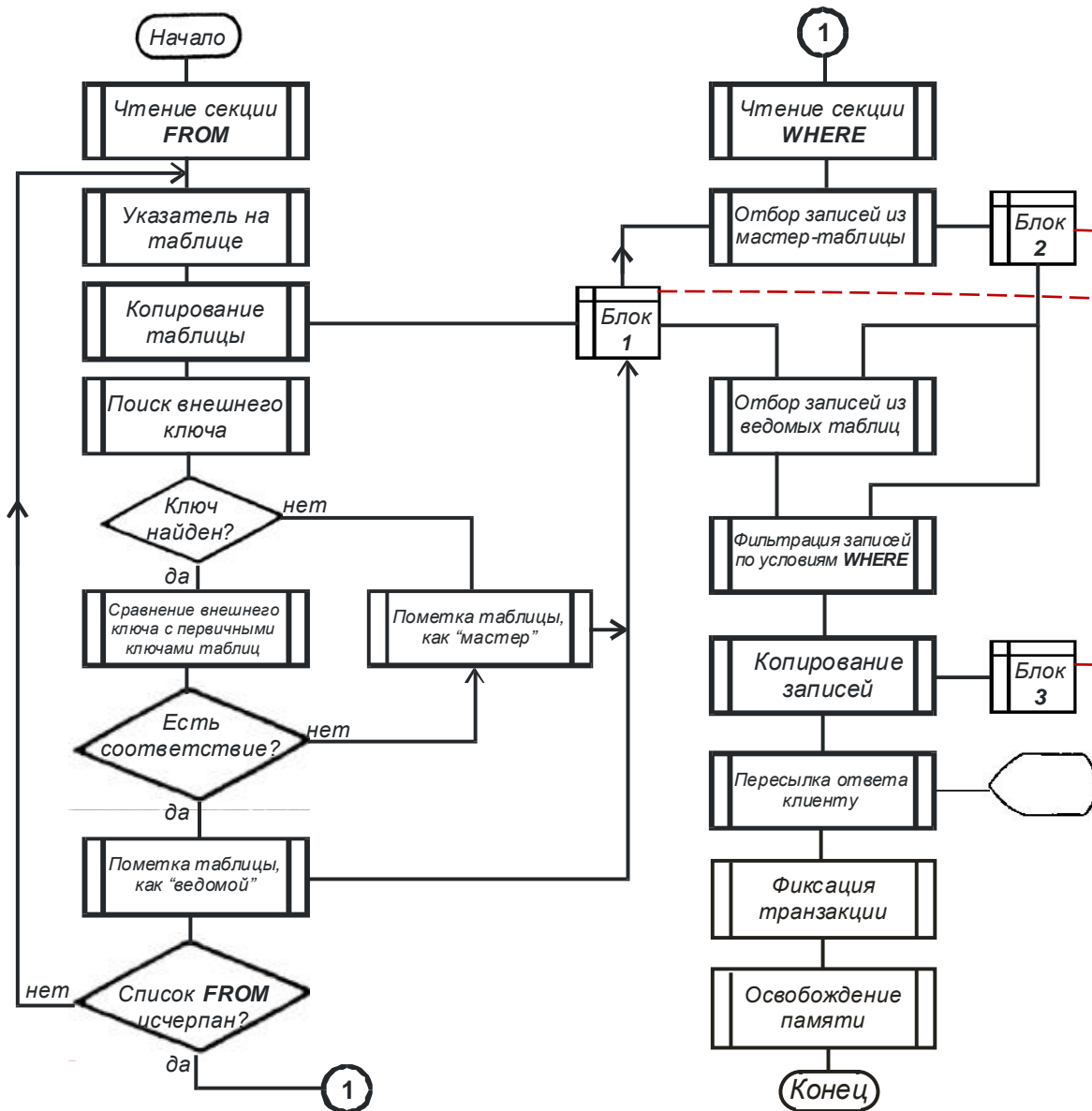
### СУБД ORACLE Database 11g R2 (12c)

Файл данных имеет предельный размер, обусловленными возможностями адресации файловой системы операционной системы сервера. Стратегия управления определяется администратором базы данных. Он может, как увеличивать размер файла данных, так и добавлять файлы данных. Максимальное количество файлов данных для одного экземпляра системы баз данных составляет 4095.

### СУБД PostgreSQL v. 14 (v. 15)

Файл данных имеет заранее определенный размер в 1 ГБайт, что обуславливает неизменность его размера в любых операционных системах. Стратегия управления файлами данных однозначно задана архитектурой СУБД и допускает только добавление новых файлов данных в директорию %PG\_HOME%\base. Максимальное количество файлов данных для одного экземпляра системы баз данных составляет 32767.

# Организация технической инфраструктуры



## Наихудший случай:

запрос осуществляет выборку данных из всех таблиц =>

*Объем блока 1 = объему всей базы данных*

условие отбора предполагает выборку всех строк из мастер-таблицы =>

*Объем блока 2 = объему всей базы данных*

по условию соединения таблиц предполагается выборка всех строк из всех дочерних таблиц =>

*Объем блока 3 = объему всей базы данных*

Для осуществления успешной выборки  
(без использования файла подкачки ОС):

$$V_{\text{ГСО}} = V_{\text{блок1}} + V_{\text{блок2}} + V_{\text{блок3}} = 3 \times V_{\text{БД}}$$



# Литература

1. **Кайт, Т., Кун, Д.** Oracle для профессионалов: архитектура и методики программирования, 3-е изд.: Пер. с англ. – Москва: ООО "ИД Вильямс", 2016. – 960 с.
2. **Дейт, К. Дж.** Введение в системы баз данных: пер. с англ. – 7-е изд. / К. Дж. Дейт. – Москва: Издательский дом "Вильямс", 2001. – 1072 с.
3. **Кузнецов, С. Д.** Основы баз данных: курс лекций : учеб. пособие для студентов вузов, обучающихся по специальностям в области информационных технологий / С. Д. Кузнецов. – Москва: Интернет-ун-т ин-форм. технологий, 2005. – 488 с.
4. **Бойко, В. В.,** Проектирование баз данных информационных систем / В. В. Бойко, В. М. Савинков. – Москва: Финансы и статистика, 1989.
5. **Смирнов С. Н. Задворьев И.С.** Работаем с ORACLE.: Учебное пособие/2-е изд., испр. и доп. – М: Гелиос АРВ, 2002 г. – 496 с.
6. **ISO/TR 16044:** 2004 – Graphic technology – Database architecture model and control parameter coding for process control and workflow (Database AMPAC).
7. **Кормен, Т.** Алгоритмы: построение и анализ / Т. Кормен, Ч. Л. Лейзерсон, Р. Ривест. – Москва: МЦНМО, 1999. – 960 с.
8. **Задворьев, И. С.,** Язык PL/SQL. Учебно-методическое пособие. – Москва: Онто-Принт, 2017. – 178 с.
9. **Тараканов О.В.,** Инженерия систем баз данных. Уч. Пособие. – Орел: Академия ФСО России, 2021. – 392 с.
10. **ISO/TR 16044:** 2004 – Graphic technology – Database architecture model and control parameter coding for process control and workflow (Database AMPAC).

Электронные ресурсы образовательного портала [ACADEMY.ORACLE.COM](https://academy.oracle.com).

Электронные ресурсы образовательного портала [INTUIT.RU](https://intuit.ru).

Электронные ресурсы портала [HTTPS://ORACLEPLUSQL.RU](https://oracleplusql.ru).

Электронные ресурсы портала [POSTGRESPRO.RU](https://postgrespro.ru).