


**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

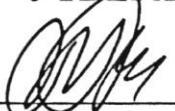
Департамент анализа данных, принятия решений и финансовых технологий

СОГЛАСОВАНО

ООО «Зеробит»
Руководитель отдела по работе с
корпоративными заказчиками
 К.Г.Базанов

«19» ноября 2019 г.

УТВЕРЖДАЮ

 Ректор
М.А. Эскиндаров

«16» ноября 2019 г.

Шевцов В.В.

ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся
по направлению подготовки 09.03.03 «Прикладная информатика»,
профили «ИТ-сервисы и технологии обработки данных в экономике и
финансах», «Высокопроизводительные вычисления в цифровой
экономике»

*Рекомендовано Ученым советом
Факультета прикладной математики и информационных технологий
(протокол №19 от 19.11.2019 г.)*

*Одобрено Советом учебно-научного Департамента анализа данных,
принятия решений и финансовых технологий
(протокол №5 от 31.10.2019 г.)*

Москва 2019

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

Департамент анализа данных, принятия решений и финансовых технологий

Шевцов В.В.

ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся
по направлению подготовки 09.03.03 «Прикладная информатика»,
профили «ИТ-сервисы и технологии обработки данных в экономике и
финансах», «Высокопроизводительные вычисления в цифровой
экономике»

Москва 2019

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

Департамент анализа данных, принятия решений и финансовых технологий

СОГЛАСОВАНО

УТВЕРЖДАЮ

ООО «Зеробит»
Руководитель отдела по работе с
корпоративными заказчиками
_____ К.Г.Базанов

Ректор

_____ М.А. Эскиндаров

19.11.2019 г.

26.11.2019 г.

Шевцов В.В.

ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся
по направлению подготовки 09.03.03 «Прикладная информатика»,
профили «ИТ-сервисы и технологии обработки данных в экономике и
финансах», «Высокопроизводительные вычисления в цифровой
экономике»

*Рекомендовано Ученым советом
Факультета прикладной математики и информационных технологий
(протокол №19 от 19.11.2019 г.)*

*Одобрено Советом учебно-научного Департамента анализа данных,
принятия решений и финансовых технологий
(протокол №5 от 31.10.2019 г.)*

Москва 2019

УДК519(073)
ББК22.16я73
Б95

Рецензент: **Феклин В.Г.**, кандидат физико-математических наук,
доцент Департамента анализа данных, принятия
решений и финансовых технологий Финансового
университета

Шевцов В.В.

«Технологии обработки больших данных». Рабочая программа дисциплины
для студентов, обучающихся по направлению подготовки 09.03.03
«Прикладная информатика», профили «ИТ-сервисы и технологии обработки
данных в экономике и финансах», «Высокопроизводительные вычисления в
цифровой экономике» – М.: Финансовый университет, Департамент анализа
данных, принятия решений и финансовых технологий, 2019. – 23 с.

Дисциплина «Технологии обработки больших данных» относится к Модулю
профиля «ИТ-сервисы и технологии обработки данных в экономике и финансах»,
направления подготовки 09.03.03 «Прикладная информатика».

Дисциплина «Технологии обработки больших данных» относится к
Профильному блоку дисциплин по выбору (для 2018 года приема), к Модулю
дисциплин по выбору, углубляющих освоение профиля «Высокопроизводительные
вычисления в цифровой экономике» (для 2019, 2020 год приема), направление
подготовки 09.03.03 «Прикладная информатика».

Рабочая программа включает требования к освоению дисциплины, тематику
лекций и практических занятий, учебно-методическое и информационное
обеспечение.

УДК
519(073)
ББК
22.16я73
Б95

Учебное издание

Шевцов Василий Викторович

Технологии обработки больших данных

Рабочая программа дисциплины

Компьютерный набор, верстка: Шевцов В.В.

Формат 60×90/16. Гарнитура *Times New Roman*

Усл. п. л. ____ Изд. № ____ Тираж ____

Заказ № _____

© **Шевцов В.В., 2019**

©

© **Финансовый университет, 2019**

Оглавление

1. Наименование дисциплины	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	2
3. Место дисциплины в структуре образовательных программ	5
4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся (в семестре, в сессию)	7
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	9
5.1. Содержание дисциплины	9
5.2. Учебно–тематический план	10
5.3. Содержание семинаров, практических занятий	122
6. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине	144
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	144
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	155
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	155
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины:	22
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	22
10. Методические указания для обучающихся по освоению дисциплины	232
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	233
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	243

1. Наименование дисциплины

«Технологии обработки больших данных».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Профиль «ИТ- сервисы и технологии обработки данных в экономике и финансах»

Код компет енции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
<i>2017, 2018 год приема</i>			
ПК-6	Способность собирать детальную информацию для формализации требований пользователей заказчика	-	<u>Знать</u> : современные инструменты и методы управления организацией, в том числе методы планирования деятельности, распределение поручений, контроля исполнения принятия решений; программные средства ИТ организации; методологию и инструментарий сбора информации о информационных потребностях пользователей; <u>Уметь</u> : использовать, обобщать и анализировать информацию, ставить цели и находить пути их достижения в условиях формирования и развития информационного общества; разрабатывать программные средства; реализовывать базовые методы сбора информации о информационных потребностях пользователей; <u>Владеть</u> : способностью собирать детальную информацию для формализации требований пользователей заказчика, инструментами и методами моделирования информационных процессов предприятия, навыками сбора детальную информации и формализации требований пользователей.
ПК-7	Способность проводить описание прикладных процессов и информационного	-	<u>Знать</u> : основные этапы проектирования информационных систем, основанных на объектном подходе с использованием промышленных стандартизированных решений; этапы и модели жизненного

	обеспечения решения прикладных задач		<p>цикла программах изделий; современные методы и технологии моделирования бизнес- процессов.</p> <p>Уметь: определять требования к проектируемой информационной системе; моделировать информационные и прикладные (бизнес) процессы; описывать программное изделие; разрабатывать руководство пользователя к программе; применять инструментарий описания бизнес- процессов</p> <p>Владеть: способностью проводить описание прикладных процессов и информационного обеспечения решения прикладных задач методами, средствами и технологией анализа информационных ресурсов предметных областей, методиками проектирования, тестирования и отладки программных продуктов в среде быстрой разработки приложений, навыками применения современных программных средств моделирования бизнес-процессов.</p>
ПК-13	Способность осуществлять установку и настройку параметров программного обеспечения информационных систем	-	<p>Знать: основные компоненты программного обеспечения</p> <p>Уметь: производить установку и настройку программного обеспечения</p> <p>Владеть: сформировавшимися систематическими знаниями основных компонентов программного обеспечения; сформировавшимися систематическим умением производить установку и настройку программного обеспечения.</p>
ПК-14	Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	-	<p>Знать: модели данных</p> <p>Уметь: разрабатывать концептуальную модель прикладной области, выбирать инструментальные средства и технологии проектирования ИС</p> <p>Владеть: сформировавшимися систематическими знаниями моделей данных; сформировавшимся систематическим умением разрабатывать концептуальную модель прикладной области, выбирать инструментальные средства и технологии проектирования ИС</p>

ПКП-2	Способность применять методы внедрения и эксплуатации корпоративных информационных систем в сфере экономики и финансов	-	<p><u>Знать:</u> принципы работы облачных платформ в области обработки больших данных, правила доступа к ним, работы, источники больших данных, методы внедрения и эксплуатации корпоративных информационных систем</p> <p><u>Уметь:</u> вводить в эксплуатацию корпоративные информационные системы в сфере экономики и финансов, находить, визуализировать и обрабатывать данные</p> <p><u>Владеть:</u> принципами работы реляционных СУБД, NoSQL, NotOnlySQL, принципами работы распределенных файловых систем и систем обработки данных, основными инструментами для поиска и сбора необходимой информации</p>
ПКП-4	Способность применять технологии моделирования и анализа процессов в сфере экономики и финансов	-	<p><u>Знать:</u> языки программирования, команды, методы и методики работы с большими данными</p> <p><u>Уметь:</u> Выбрать платформу для работы с распределенными данными в зависимости от решаемой задачи, анализировать эффективность работы с большими данными, проводить оценки полученным результатам</p> <p><u>Владеть:</u> методами и методиками анализа полученных результатов</p>

2019, 2020 год приема

ПКП-2	Способность применять методы внедрения и эксплуатации корпоративных информационных систем в сфере экономики и финансов	<p>1. Демонстрирует знание назначения и функционал типовых модулей корпоративных информационных систем, основные этапы и методы их внедрения и эксплуатации.</p> <p>2. Владеет методологией внедрения и эксплуатации приложений в сфере экономики и финансов на платформе корпоративных информационных систем.</p>	<p><u>Знать:</u> назначение и функционал типовых модулей корпоративных информационных систем</p> <p><u>Уметь:</u> внедрять и эксплуатировать типовые модули корпоративных информационных систем</p> <p><u>Знать:</u> методологию внедрения и эксплуатации приложений в сфере экономики и финансов</p> <p><u>Уметь:</u> внедрять и эксплуатировать приложения на платформе корпоративных информационных систем.</p>
-------	--	--	--

ПКП-3	Способность применять методы разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем	<p>1. Демонстрирует знание назначения и функционал типовых модулей корпоративных информационных систем, основные методы разработки приложений на их платформе</p> <p>2. Владеет методологией разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем</p>	<p>Знать: назначение и функционал типовых модулей корпоративных информационных систем</p> <p>Уметь: разрабатывать приложения на платформе типовых модулей корпоративных информационных систем</p> <p>Знать состав экосистем в области обработки больших данных.</p> <p>Уметь работать с отдельными компонентами экосистем в области обработки больших данных</p>
-------	--	--	--

Профиль «Высокопроизводительные вычисления в цифровой экономике»

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
2018 год приема			
ПКП-7	Способность провести статистический анализ данных, в том числе разнородных и данных большого объема	-	<p>Знать: методы анализа статистических данных, методы и методики работы с большими данными</p> <p>Уметь: рассчитывать показатели качества статистического анализа данных в зависимости от объема выборки, анализировать эффективность работы с большими данными</p> <p>Владеть: языками программирования, методами и методиками анализа полученных результатов</p>
2019, 2020 год приема			
ПКП-5	Способность принимать участие в реализации профессиональных коммуникаций в рамках разработки междисциплинарных	1. Анализирует информацию, полученную от специалистов в различных областях экономики для формализации задачи	<p>Знать: методы анализа статистических данных, методы и методики работы с большими данными</p> <p>Уметь: рассчитывать показатели качества статистического анализа данных в зависимости от объема выборки, анализировать</p>

	ых проектов	2.Объясняет особенности алгоритмического обеспечения решения задачи для неспециалистов в области IT-технологий.	<p>эффективность работы с большими данными</p> <p>Знать: особенности алгоритмического обеспечения решения профессиональной задачи</p> <p>Уметь: объяснять особенности алгоритмического решения задачи неспециалистам в области IT-технологий.</p>
ПКП-7	Способность провести статистический анализ данных, в том числе разнородных и данных большого объема	<p>1. Демонстрирует методы анализа статистических данных, рассчитывает показатели качества статистического анализа данных в зависимости от объема выборки.</p> <p>2. Разрабатывает алгоритмы для обработки больших данных.</p> <p>3. Обосновывает выбор методов статистического анализа в зависимости от типов данных, в том числе нечисловых данных.</p>	<p>Знать: методы анализа статистических данных</p> <p>Уметь: рассчитывать показатели качества статистического анализа данных в зависимости от объема выборки</p> <p>Знать: языки программирования, команды, методы и методики работы с большими данными</p> <p>Уметь: анализировать эффективность работы с большими данными</p> <p>Знать: методы и методики анализа полученных результатов</p> <p>Уметь: проводить оценки полученным результатам</p>

3. Место дисциплины в структуре образовательных программ

Дисциплина «Технологии обработки больших данных» относится к Модулю профиля «ИТ-сервисы и технологии обработки данных в экономике и финансах», направления подготовки 09.03.03 «Прикладная информатика».

Дисциплина «Технологии обработки больших данных» относится к Профильному блоку дисциплин по выбору (для 2018 года приема), к Модулю дисциплин по выбору, углубляющих освоение профиля «Высокопроизводительные вычисления в цифровой экономике» (для 2019, 2020 год приема), направление подготовки 09.03.03 «Прикладная информатика».

4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся (в семестре, в сессию)

Профиль: "ИТ-сервисы и технологии обработки данных в экономике и финансах"

Общая трудоёмкость дисциплины составляет 4 зачётные единицы

Вид промежуточной аттестации – экзамен

Вид текущего контроля – контрольная работа

Очная форма обучения 2017/2018 г.г.

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоёмкость дисциплины	4/144	144
Контактная работа – Аудиторные занятия	54/50	54/50
<i>Лекции</i>	18/16	18/16
<i>Семинары, практические занятия</i>	36/34	36/34
Самостоятельная работа	90/94	90/94
Вид текущего контроля	Экзамен	Экзамен
Вид промежуточной аттестации	Контрольная работа	Контрольная работа

Общая трудоёмкость дисциплины составляет 5 зачётных единиц

Вид промежуточной аттестации – экзамен

Вид текущего контроля – контрольная работа

Очная форма обучения 2019, 2020 г.г.

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 5 (в часах)
Общая трудоёмкость дисциплины	5/180	180
Контактная работа – Аудиторные занятия	68	68
<i>Лекции</i>	34	34
<i>Семинары, практические занятия</i>	34	34
Самостоятельная работа	112	112
Вид текущего контроля	Экзамен	Экзамен
Вид промежуточной аттестации	Контрольная работа	Контрольная работа

Очно-заочная форма обучения 2019/2020 г.г.

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 5/4 (в часах)
Общая трудоемкость дисциплины	5/180	180
Контактная работа – Аудиторные занятия	36	36
<i>Лекции</i>	18	18
<i>Семинары, практические занятия</i>	18	18
Самостоятельная работа	144	144
Вид текущего контроля	Экзамен	Экзамен
Вид промежуточной аттестации	Контрольная работа	Контрольная работа

Профиль: "Высокопроизводительные вычисления в цифровой экономике"

Общая трудоёмкость дисциплины составляет 4 зачётные единицы

Вид промежуточной аттестации – экзамен

Вид текущего контроля – контрольная работа

Заочная форма обучения, 2018 г.

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 9 (в часах)
Общая трудоемкость дисциплины	4/144	144
Контактная работа – Аудиторные занятия	16	16
<i>Лекции</i>	4	4
<i>Семинары, практические занятия</i>	12	12
Самостоятельная работа	128	128
Вид текущего контроля	Экзамен	Экзамен
Вид промежуточной аттестации	Контрольная работа	Контрольная работа

Общая трудоёмкость дисциплины составляет 4 зачётные единицы

Вид промежуточной аттестации – зачет

Вид текущего контроля – контрольная работа

Заочная форма обучения 2019, 2020 г.г.

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 9 (в часах)
Общая трудоемкость дисциплины	4/144	144
Контактная работа – Аудиторные занятия	16	16
<i>Лекции</i>	4	4

<i>Семинары, практические занятия</i>	<i>12</i>	<i>12</i>
<i>Самостоятельная работа</i>	<i>128</i>	<i>128</i>
Вид текущего контроля	Экзамен	Экзамен
Вид промежуточной аттестации	Контрольная работа	Контрольная работа

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

1. Тема: Документно-ориентированная СУБД MongoDB

- 1.1. Историческая справка
- 1.2. SQL, NoSQL, NotOnlySQL
- 1.3. Определения документно-ориентированной СУБД
- 1.4. Состав и операции MongoDB

2. Тема: Запросы

- 2.1. Принцип построение запросов, сравнение с SQL
- 2.2. Селекторы запроса
- 2.3. Дополнительные средства

3. Тема: Агрегирование

- 3.1. Минимум и максимум
- 3.2. Команда distinct
- 3.3. Команда group
- 3.4. Map-reduce

4. Тема: Обновление и удаление

- 4.1. Типы и параметры операций обновления
- 4.2. Операторы обновления
- 4.3. Атомарная обработка документа
- 4.4. Операции удаления
- 4.5. Параллелизм, атомарность и изолированность

5. Тема: Подключение к приложению

- 5.1. Подключение из JavaScript
- 5.2. Подключение из PHP
- 5.3. Подключение из C++
- 5.4. Подключение из MS SQL Server

6. Тема: Индексирование и оптимизация запросов

- 6.1. Основные понятия индексирования
- 6.2. B-деревья
- 6.3. Типы индексов
- 6.4. Администрирование индексов
- 6.5. Оптимизация: выявление и исследование медленных запросов

7. Тема: Репликация и сегментирование

- 7.1. Назначение репликации
- 7.2. Сценарии репликации
- 7.3. Настройка и администрирование
- 7.4. Драйверы и репликация
- 7.5. Общие понятия сегментирования
- 7.6. Сегментированный кластер и запросы к нему
- 7.7. Сегментный ключ
- 7.8. Администрирование сегментированного кластера

8. Тема: Развертывание и администрирование

- 8.1. Развертывание
- 8.2. Конфигурирование сервера
- 8.3. Импорт и экспорт данных
- 8.4. Безопасность
- 8.5. Мониторинг и диагностика
- 8.6. Резервное копирование и восстановление
- 8.7. Производительность

5.2. Учебно–тематический план

Профиль: "ИТ-сервисы и технологии обработки данных в экономике и финансах"

Очная форма обучения 2017/2018 г.г.

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах						Формы текущего контроля успеваемости и
		Все го	Аудиторная работа				Самос тоятел ьная работа	
			Общ ая, в т.ч.:	Лекции	Семина ры, практи ческие заняти я	Занятия в интеракти вных формах		
1.	Документно-ориентированная СУБД MongoDB	22/18	12/8	4/2	8/6	4/2	10/10	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2.	Запросы	16/18	6/6	2/2	4/4	2/2	10/12	
3.	Агрегирование	16/18	6/6	2/2	4/4	2/2	10/12	
4.	Обновление и удаление	18/18	6/6	2/2	4/4	2/2	12/12	
5.	Подключение к приложению	18/18	6/6	2/2	4/4	2/2	12/12	
6.	Индексирование и оптимизация запросов	18/18	6/6	2/2	4/4	2/2	12/12	

7.	Репликация и сегментирование	18/18	6/6	2/2	4/4	2/2	12/12	
8.	Развертывание и администрирование	18/18	6/6	2/2	4/4	2/2	12/12	
	В целом по дисциплине	144	54/50	18/16	36/34	18/16	90/94	Контрольная работа
	Итого в %					33/32%		

Очная / очно-заочная форма обучения 2019, 2020 г.г.

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах						Формы текущего контроля успеваемости
		Все го	Аудиторная работа				Самост оятельн ая работа	
			Общ ая, в т.ч.:	Лекции	Семи на ры, практ ическ ие занят ия	Занятия в интеракт ивных формах		
1.	Документно-ориентированная СУБД MongoDB	22	8/4	4/2	4/2	2	14/18	Самостоятель ные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2.	Запросы	26	12/8	6/4	6/4	2	14/18	
3.	Агрегирование	22	8/4	4/2	4/2	2	14/18	
4.	Обновление и удаление	22	8/4	4/2	4/2	2	14/18	
5.	Подключение к приложению	22	8/4	4/2	4/2	2	14/18	
6.	Индексирование и оптимизация запросов	22	8/4	4/2	4/2	2	14/18	
7.	Репликация и сегментирование	22	8/4	4/2	4/2	2	14/18	
8.	Развертывание и администрирование	22	8/4	4/2	4/2	2	14/18	
	В целом по дисциплине	180	68/36	34/18	34/18	16	112/144	Контрольная работа
	Итого в %					23/50		

Профиль: "Высокопроизводительные вычисления в цифровой экономике"

Заочная форма обучения 2018, 2019, 2020 г.г.

№ п/п	Наименование тем (разделов)	Трудоемкость в часах			Формы текущего
		Все	Аудиторная работа	Самост	

	дисциплины	го	Общ ая, в т.ч.:	Лекц ии	Семина ры, практич еские занятия	Занятия в интеракт ивных формах	оательн ая работа	контроля успеваемости
1.	Документно-ориентированная СУБД MongoDB	20	4	2	2	1	16	Самостоятельн ые работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2.	Запросы	18	2	-	2	1	16	
3.	Агрегирование	17	1	-	1	1	16	
4.	Обновление и удаление	19	3	2	1	1	16	
5.	Подключение к приложению	18	2	-	2	1	16	
6.	Индексирование и оптимизация запросов	17	1	-	1	1	16	
7.	Репликация и сегментирование	17	1	-	1	1	16	
8.	Развертывание и администрирование	18	2	-	2	1	16	
	В целом по дисциплине	144	16	4	12	8	128	Контрольная работа
	Итого в %					50%		

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Документно-ориентированная СУБД MongoDB	Программные оболочки Подключение к серверу MongoDB Создание объектов MongoDB Создание документов Удаление документов и объектов MongoDB <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Запросы	Создание запросов к БД <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

Агрегирование	Использование функции агрегирования в запросах БД <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Обновление и удаление	Создание операций на обновление и удаление данных <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Подключение к приложению	Отработка подключений из различных сред программирования Создание пользовательского Web-интерфейса Подключение MongoDB к приложению и настройка взаимодействия <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Индексирование и оптимизация запросов	Создание и управление индексами Выявление и исследование медленных запросов Оптимизация запросов <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Репликация и сегментирование	Настройка и администрирование репликации Настройка и администрирование сегментирования <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Развертывание и администрирование	Создание серверов MongoDB в виртуальной среде и их настройка Отработка операций импорта и экспорта данных Создание пользователей и ролей Отработка операций резервного копирования и восстановления Отработка операций мониторинга производительности <i>Рекомендуемые источники: п.8, [1], [2]</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование разделов, тем входящих в дисциплину	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Документно-ориентированная СУБД MongoDB	Установка СУБД, создание базы. Создание, обновление и удаление документов. Выборка данных из коллекций.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Запросы	Особенности создание запросов к БД	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Агрегирование	Применение функции агрегирования в запросах БД.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Обновление и удаление	Работа с документами созданной базы данных, освоение методов выборки данных из коллекции MongoDB	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Подключение к приложению	Пользовательский Web-интерфейса Настройка взаимодействия приложения и MongoDB.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Индексирование и оптимизация запросов	Работа с индексами. Определение медленных запросов и оптимизация запросов.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Репликация и сегментирование	Основы администрирования репликации и сегментирования	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

Развертывание и администрирование	Сервер MongoDB в виртуальной среде и его настройка Импорта и экспорта данных Создание пользователей и ролей, сохранение и восстановление баз данных.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
-----------------------------------	--	--

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный перечень вопросов к контрольной работе

1. Создание и операции с объектами MongoDB
2. Запросы на выборку данных
3. Обновление данных
4. Удаление данных
5. Агрегирование
6. Операции с индексами

Примерные задания контрольной работы

1. Назовите различия noSQL-баз данных и реляционных баз данных
2. Перечислите методы обработки данных в MondoDB, схожие с реляционными БД.
3. Напишите 3 примера операции агрегации данных в MondoDB.
4. Назовите сходства noSQL-баз данных и реляционных баз данных
5. Перечислите 10 методов работы с курсорами в MongoDB и дайте им описание.
6. Напишите пример формирования курсора в MongoDB, включая: условия, сортировку, набор выводимых ключей.
7. Напишите пример формирования курсора в MongoDB, включая: условия, сортировку, набор выводимых ключей и применение метода подсчета количества элементов.

Критерии балльной оценки различных форм текущего контроля успеваемости

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных, принятия решений и финансовых технологий.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине:

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для
оценки индикаторов достижения компетенций, умений и знаний**

**Профиль «ИТ – сервисы и технологии обработки данных в экономике и
финансах»**

2019, 2020 год приема

Код компетенции	Наименование компетенции	Примеры заданий для оценки индикаторов достижения компетенций
ПКП-2	Способность применять методы внедрения и эксплуатации корпоративных информационных систем в сфере экономики и финансов	<p>1.Демонстрирует знание назначения и функционал типовых модулей корпоративных информационных систем, основные этапы и методы их внедрения и эксплуатации.</p> <p align="center">Задание 1.</p> <p>Перечислите методы обработки данных в MondoDB, схожие с реляционными БД</p> <p>2.Владеет методологией внедрения и эксплуатации приложений в сфере экономики и финансов на платформе корпоративных информационных систем.</p> <p align="center">Задание 2.</p> <p>Создайте резервную копию своей рабочей базы данных и восстановите в новую базу данных резервную копию вашей рабочей базы данных.</p>
ПКП-3	Способность применять методы разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем	<p>1.Демонстрирует знание назначения и функционал типовых модулей корпоративных информационных систем, основные методы разработки приложений на их платформе</p> <p align="center">Задание 1.</p> <p>Напишите 3 примера операции агрегации данных в MondoDB.</p> <p>2.Владеет методологией разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем</p> <p align="center">Задание 2.</p> <p>Напишите 5 примеров с командами HDFS для работы с файлами</p>

2017, 2018 год приема

Код компетенции	Наименование компетенции	Примеры заданий для оценки сформированности компетенции
ПК-6	Способность собирать детальную информацию для	<p align="center">Задание 1.</p> <p>Описание данных: mails.json Переписка компании Enron</p>

	формализации требований пользователей заказчика	<p>Для каждого дня недели посчитайте, сколько писем было отправлено на адрес ebass@enron.com?</p> <p>Покопаемся в переписке Shanna Husser и Eric Bass. Сколько писем каждый из них отправил другому? Laurie Ellis иногда посылает письма с одинаковыми темами. Для каждой темы письма, использованной в 2000 году, посчитайте, сколько раз она была использована.</p> <p>Сколько человек отправляют письма сами себе?</p> <p>В какой папке больше всего писем?</p> <p>Сколько различных отправителей в датасете?</p> <p>В какой день недели отправлено максимальное количество писем?</p>
ПК-7	Способность проводить описание прикладных процессов и информационного обеспечения решения прикладных задач	<p>Задание 1.</p> <p>Создайте курсор, содержащий 10 записей действующих студентов Экологического факультета, которые находятся с 90-й записи в списке. Выведите в цикле номера студенческих билетов.</p> <p>Создайте цикл, выводящий номер студенческого билета и порядковый номер студента в списке.</p> <p>Ограничьте количество выводимых записей.</p> <p>Создать код для пагинации списка студентов.</p> <p>Задавайте в переменные номер страницы и количество объектов на странице. Запрос должен выводить содержимое заданной страницы (Количество выводимых полей – не более 5. Состав полей – на Ваше усмотрение.)</p>
ПК-13	Способность осуществлять установку и настройку параметров программного обеспечения информационных систем	<p>Задание 1.</p> <p>Загрузите файл JSON по ссылке: mongodb.json</p> <p>На локальном компьютере сохраните загруженный файл [JSON] в каталог, к которому вы сможете получить доступ позже в этой лаборатории.</p> <p>Загрузите MongoDB Community Server edition по этой ссылке: https://www.mongodb.com/download-center?jmp=nav#community</p> <p>После загрузки установите MongoDB Community Edition на локальном компьютере.</p> <p>После завершения установки перейдите в папку установки на локальном компьютере.</p>
ПК-14	Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	<p>Описание данных: kladr01.json-kladr08.json - Классификатор адресов Российской Федерации. Данные описывают адреса вплоть до населенных пунктов.</p> <p>Импортируйте КЛАДР в локальный MongoDB.</p> <p>Сколько уровней адресов в КЛАДРе?</p> <p>Перечислите наименование уровней.</p> <p>Сколько объектов в каждом уровне?</p> <p>Составьте перечень типов населенных пунктов.</p> <p>Сколько населенных пунктов начинается на букву М?</p>
ПКП-2	Способность применять методы внедрения и	<p>Задание 1.</p> <p>Назовите различия поSQL-баз данных и реляционных баз данных</p>

	эксплуатации корпоративных информационных систем в сфере экономики и финансов	<p>Назовите сходства noSQL-баз данных и реляционных баз данных</p> <p>Задание 2. Перечислите методы обработки данных в MondoDB, схожие с реляционными БД Опишите принципы обработки информации в соответствии с парадигмой MapReduce.</p> <p>Задание 3. Создайте резервную копию своей рабочей базы данных. Восстановите в новую базу данных резервную копию вашей рабочей базы данных. Восстановите в новую базу данных резервную копию одной коллекции вашей рабочей базы данных.</p>
ПКП-4	Способность применять технологии моделирования и анализа процессов в сфере экономики и финансов	<p>Задание 1. Напишите 3 примера операции агрегации данных в MondoDB.</p> <p>Задание 2. Напишите пример формирования курсора в MongoDB, включая: условия, сортировку, набор выводимых ключей. Создайте уникальный индекс по ключу. Создайте уникальный составной индекс по ключам</p> <p>Задание 3. Напишите 5 примеров с командами HDFS для работы с файлами Перечислите 10 методов работы с курсорами в MongoDB и дайте им описание.</p>

Профиль «Высокопроизводительные вычисления в цифровой экономике»

2019, 2020 год приема

Код компетенции	Наименование компетенции	Примеры заданий для оценки индикаторов компетенций
ПКП-5	Способность принимать участие в реализации профессиональных коммуникаций в рамках разработки междисциплинарных проектов	<p>1.Анализирует информацию, полученную от специалистов в различных областях экономики для формализации задачи</p> <p>Задание 1. Описание данных: kladr01.json-kladr08.json - Классификатор адресов Российской Федерации. Данные описывают адреса вплоть до населенных пунктов. Импортируйте КЛАДР в локальный MongoDB. Сколько уровней адресов в КЛАДРе?</p> <p>2.Объясняет особенности алгоритмического обеспечения решения задачи для неспециалистов в области IT-технологий.</p>

		<p>Задание 2. Назовите различия и сходства noSQL-баз данных и реляционных баз данных.</p>
ПКП-7	Способность провести статистический анализ данных, в том числе разнородных и данных большого объема	<p>1. Демонстрирует методы анализа статистических данных, рассчитывает показатели качества статистического анализа данных в зависимости от объема выборки.</p> <p>Задание 1. Описание данных: mails.json Переписка некоторой компании. Для каждого дня недели посчитайте, сколько писем было отправлено на адрес ebass@enron.com? Проанализируем переписку Shanna Husser и Eric Bass. Сколько писем каждый из них отправил другому? Laurie Ellis иногда посылает письма с одинаковыми темами. Для каждой темы письма, использованной в 2000 году, посчитайте, сколько раз она была использована. Сколько человек отправляют письма сами себе? В какой папке меньше всего писем? Сколько различных отправителей в датасете?</p> <p>2. Разрабатывает алгоритмы для обработки больших данных.</p> <p>Задание 2. Перечислите методы обработки данных в MondoDB, схожие с реляционными БД Перечислите 10 методов работы с курсорами в MongoDB и дайте им описание.</p> <p>3. Обосновывает выбор методов статистического анализа в зависимости от типов данных, в том числе нечисловых данных.</p> <p>Задание 3. Создайте курсор, содержащий 10 записей действующих студентов Экологического факультета, которые находятся с 90-й записи в списке. Выведите в цикле номера студенческих билетов. Создать код для пагинации списка студентов. Задавайте в переменные номер страницы и количество объектов на странице. Запрос должен выводить содержимое заданной страницы.</p>

Примеры практико-ориентированных (ситуационных) заданий

1. Импортируйте КЛАДР в локальный MongoDB.
2. Сколько уровней адресов в КЛАДРе?
3. Перечислите наименование уровней.
4. Сколько объектов в каждом уровне?

5. Составьте перечень типов населенных пунктов.
6. Сколько населенных пунктов начинается на букву М?
7. Сколько населенных пунктов начинается на каждую букву?
8. Какое назначение ключа district?
9. Какое количество объектов у каждого значения district?
10. Добавление массива
 - 10.1. Создайте курсор: первые 100 документов населенных пунктов 4 уровня. Ключи документов: `_id`, `type`, `name`.
 - 10.2. Добавьте курсор в новую коллекцию. Выведите результат добавления.
 - 10.3. Преобразуйте курсор в массив.
 - 10.4. Добавьте массив в коллекцию. Выведите результат добавления.
 - 10.5. Вставьте в новую коллекцию с 200 по 300 документов населенных пунктов 4 уровня, входящих в Центральный федеральный округ.
11. Создайте индекс по ключу `wikiname`.
12. Создайте уникальный индекс по ключу `name` населенного пункта. Какой будет результат?
13. Создайте уникальный составной индекс по ключам `name` и `code` населенного пункта. Какой будет результат?
14. Продемонстрируйте на любом курсоре работу функции `showRecordId()`
15. Создайте резервную копию своей рабочей базы данных.
16. Восстановите в новую базу данных резервную копию вашей рабочей базы данных.
17. Восстановите в новую базу данных резервную копию одной коллекции вашей рабочей базы данных.

Примерные вопросы для подготовки к зачету / экзамену

1. Назовите различия noSQL-баз данных и реляционных баз данных
2. Перечислите типы данных MongoDB, опишите особенности их применения
3. Опишите индексы в MondoDB: общие положения, типы, особенности применения, администрирование.
4. Перечислите методы обработки данных в MondoDB, схожие с реляционными БД.
5. Дайте описание модели устройства базы данных MondoDB: наименование, назначение, функции работы с объектом.
6. Перечислите команды для работы с географическими объектами в MongoDB: синтаксис, назначение, описание параметров.
7. Профилирование реализации алгоритмов на Python, принципы решения задачи оптимизации производительности алгоритма

8. Проблема Global Interpreter Lock в Python и способы обхода ее ограничений
9. Технологический стек Python для обработки и анализа данных, Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python
10. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
11. Универсальные функции и применение функций по осям в NumPy
12. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
13. Маскирование и прихотливое индексирование в NumPy
14. Модуль multiprocessing – назначение и основные возможности, API multiprocessing.Pool
15. Различия между потоками и процессами, различие между различными планировщиками в Dask
16. Граф зависимостей задач – суть структуры данных, ее построение и использование в Dask
17. Три ключевых структуры данных Dask: их специфика и принцип выбора структуры данных при решении задач
18. Dask.Array – структура данных, специфика реализации и применения, процедура создания
19. Dask.Array – поддерживаемые операции и отличия от NumPy ndarray
20. Распараллеливание алгоритмов с помощью dask.delayed – принцип и примеры использования
21. Дополнительные параметры декоратора dask.delayed – назначение и примеры использования
22. Использование dask.delayed для объектов и операции над объектами dask.delayed, включая ограничения их использования
23. Dask.DataFrame - структура данных, специфика реализации и применения, процедура создания Dask.DataFrame
24. Ограничения использования Dask.DataFrame и операции мэппинга в Dask.DataFrame
25. Поддержка Dask.DataFrame операций работающих со скользящим окном
26. Совместное использование промежуточных результатов в Dask: принцип работы и примеры использования

27. Dask.Bag - структура данных, специфика реализации и применения, процедура создания DaskBag

28. Организация вычислений с помощью Map / Filter / Reduce : общий принцип и специфика параллельной реализации обработки данных в Dask.Bag

Пример экзаменационного билета

Билет №

1. Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение
2. Использование `dask.delayed` для объектов и операции над объектами `dask.delayed`, включая ограничения их использования
3. Продемонстрируйте на любом курсоре работу функции `showRecordId()`

8.Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Мартишин, С. А. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala : учеб. пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва : ИНФРА-М, 2019. — 235 с. + Доп. материалы. — (Высшее образование: Бакалавриат). — www.dx.doi.org/10.12737/textbook_5cc063e18baca3.52928692. - ЭБС ZNANIUM.com. —URL: <https://new.znanium.com/catalog/product/1018196> (дата обращения: 23.04.2020). - Текст : электронный
2. Форман Д. Много цифр. Анализ больших данных при помощи Excel / Д. Форман; пер. с англ. А. Соколовой – Москва: Альпина Пабли., 2016, 2017 – 461 с. –ЭБС ZNANIUM.com. ЭБС AlpinaDigital – URL: <http://znanium.com/catalog/product/551044>, (дата обращения 23.04.2020) <https://finunivers.alpinadigital.ru/book/7921> (дата обращения 23.04.2020) . – Текст : электронный

Дополнительная литература:

3. Колдаев В.Д. Структуры и алгоритмы обработки данных: учебное пособие / В.Д. Колдаев. – Москва: ИЦ РИОР: НИЦ ИНФРА-М, 2014. – 296 с. –ЭБС ZNANIUM.com. – URL: <http://znanium.com/catalog/product/418290> (дата обращения 23.04.2020) . – Текст : электронный

9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

1. <https://www.mongodb.com>
2. <https://visualstudio.microsoft.com/ru/>

3. <https://metanit.com/nosql/mongodb/>
4. <https://ru.bmstu.wiki/MongoDB>
5. <https://www.tutorialspoint.com/mongodb/>
6. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
(<http://library.fa.ru/files/elibfa.pdf>)
7. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
8. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН»
<http://biblioclub.ru/>
9. Электронно-библиотечная система Znanium <http://www.znanium.com>
10. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
11. Электронно-библиотечная система издательства «ЮРАЙТ» <https://www.biblio-online.ru/>
12. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
13. AlpinaDigital Онлайн-библиотека деловой литературы
<https://finunivers.alpinadigital.ru/>

10. Методические указания для обучающихся по освоению дисциплины

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения.

Домашние задания следует выполнять регулярно при подготовке к практическим занятиям. Контроль выполнения домашних заданий осуществляется в ходе практических занятий в процессе выборочного собеседования.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

- 11.1. Комплект лицензионного программного обеспечения:
Windows, Microsoft Office.
Антивирус ESET Endpoint Security
- 11.2. Современные профессиональные базы данных и информационные справочные системы
Информационно-правовая система «Гарант»
Информационно-правовая система «Консультант Плюс»
Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>
Система комплексного раскрытия информации «СКРИН» - <http://www.skrin.ru/>
- 11.3. Сертифицированные программные и аппаратные средства защиты информации
- не используются
- 11.4. MongoDB server
- 11.5. MongoDB Compass

11.6. Visual Studio Code или выше

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Наличие аудитории, оснащенной компьютерной техникой и проектором, с возможностью подключения к сети «Интернет».