



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА

«СОГЛАСОВАНО»

Руководитель ОП

Государственное и муниципальное управление:
социально-экономическое развитие регионов

_____ Филатов А.Ю.
(подпись) (Ф.И.О. рук. ОП)
« _____ » _____ 20 ____ г.

«УТВЕРЖДАЮ»

Директор Департамента экономических
наук

_____ Колбина Е.О.
(подпись) (Ф.И.О. зав. каф.)
« _____ » _____ 20 ____ г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Аналитика больших данных

Направление подготовки 38.04.01 Экономика

магистерская программа «Государственное и муниципальное управление: социально-
экономическое развитие регионов»

Форма подготовки очная

курс 1 семестр 2

лекции 8 час.

практические занятия 0 час.

лабораторные работы 0 час.

в том числе с использованием МАО лек. 0 час. / пр. 0 час. / лаб. 0 час.

всего часов аудиторной нагрузки 8 час.

в том числе с использованием МАО 0 час.

самостоятельная работа 100 час.

в том числе на подготовку к экзамену 0 час.

контрольные работы (количество) 0

курсовая работа / курсовой проект не предусмотрены

зачет 2 семестр

экзамен не предусмотрен

Рабочая программа составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего образования, утвержденного приказом Министерства образования и науки РФ от 12.03.2015 № 208

Рабочая программа обсуждена на заседании департамента экономических наук, протокол № _____
от « _____ » _____ 20 ____ г.

Директор департамента Е.О. Колбина

Составители: канд. физ.-мат. наук, доцент А.Ю. Филатов

ст.преп. Е.А. Воронцова

АННОТАЦИЯ

Дисциплина «Аналитика больших данных» предназначена для студентов-магистрантов первого курса, обучающихся по направлению подготовки 38.04.01 Экономика, программа «Государственное и муниципальное управление: социально-экономическое развитие регионов». Дисциплина «Аналитика больших данных» Б1.В.ДВ.02.02 входит в состав дисциплин по выбору вариативной части блока Б1

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы (108 часов). Дисциплина реализуется на 1 курсе во 2 семестре.

Учебным планом предусмотрены лекционные занятия (8 часов) и самостоятельная работа (100 часов) в соответствии с программой онлайн-курса, разработанного в СПбГУ и представленного на платформе «Открытое образование»: <https://openedu.ru/course/spbu/DTBS>.

Цель - знакомство студентов с основами обработки и анализа больших данных; изучение инструментальных средств, предназначенных для их обработки и анализа; получение практических навыков по численной обработке данных; получение практических навыков по разработке схем баз данных; изучение технологии разработки базы данных в одной из современных систем управления базами данных (от формальной постановки задачи и разработки схемы данных до создания пользовательских форм); изучение основ реляционной алгебры; изучение языка запросов SQL.

Задачи:

- дать представление жизненном цикле аналитики данных, технологиях и средствах распределенной обработки и хранения данных, базовых методах аналитики больших объемов данных, техниках визуализации данных.
- предоставить возможность использовать типовые технологии и средства аналитики данных, такие как MapReduce, Hadoop, NoSQL, язык R.
- научить обоснованно и результативно использовать, совершенствовать, разрабатывать и внедрять современные технологии и инструментальные средства анализа и работы с большими объемами данных.

В результате изучения данной дисциплины у обучающихся формируются следующие профессиональные компетенции (элементы компетенций).

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-11: способность анализировать и использовать различные источники информации для проведения экономических расчетов	Знает	современные технологии анализа больших данных
	Умеет	использовать программные продукты, в т.ч. СУБД, для анализа больших экономических данных
	Владеет	эффективными методами разработки СУБД с помощью современных технологий программирования

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Лекции (8 часов)

Тема 1. Введение в большие данные: Определение больших данных и причины их появления. Примеры возможностей для бизнеса. Различие между Business Intelligence и Big Data.

Тема 2. Жизненный цикл аналитики данных: Понятие жизненного цикла аналитики данных. Роли, необходимые для успешного создания проекта по аналитике данных.

Тема 3. Высокопроизводительные вычисления: Распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce. Проект Apache Hadoop и его экосистема. Apache Spark и его компоненты. Вычисления в реальном времени, Apache Storm, Flink.

Тема 4. Масштабирование и многоуровневое хранение данных: Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных.

Тема 5. Визуализация данных и результатов анализа: Техники визуализации данных, введение в язык R. Визуализация данных в R.

Тема 6. Сложные методы аналитики: Классификация задач анализа: Text, Data, Web, Social Mining. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация. Логистическая регрессия, ассоциации, алгоритм Априори.

Тема 7. Анализ текста: Поисковые механизмы: Lucene, Solr, ElasticSearch..Алгоритмы Word2Vec и Glove.

II. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Аналитика больших данных» представлено в Приложении 1 и включает в себя:

характеристика заданий для самостоятельной работы обучающихся и методические рекомендации по их выполнению;

требования к представлению и оформлению результатов самостоятельной работы;

критерии оценки выполнения самостоятельной работы.

III. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

Контролируемые разделы дисциплины, этапы формирования компетенций, виды оценочных средств, зачетно-экзаменационные материалы, комплекты оценочных средств для текущей аттестации, описание показателей и критериев оценивания компетенций на различных этапах их формирования, методические материалы, определяющие процедуры оценивания знаний,

умений и навыков и (или) опыта деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 2.

IV. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

1. Дьконов В. MATLAB. Полный самоучитель [Электронный ресурс]. М.: ДМК Пресс, 2014. — 768 с. — Режим доступа: <http://www.iprbookshop.ru/7911>.
2. Медведкова И.Е., Бугаев Ю.В., Чикунов С.В. Базы данных [Электронный ресурс]: учебное пособие. Воронеж: издательство Воронежского государственного университета инженерных технологий, 2014. — 108 с. — Режим доступа: <https://e.lanbook.com/book/72882>.
3. Буренин, С.Н. Web-программирование и базы данных [Электронный ресурс]: учебный практикум. М.: Московский гуманитарный университет, 2014. — Режим доступа: <http://www.iprbookshop.ru/39683>.

Дополнительная литература

(печатные и электронные издания)

1. Туманов В.Е. Основы проектирования реляционных баз данных [Электронный ресурс]. — М.: Издательство Интернет-Университета Информационных Технологий, 2014. — 224 с. — Режим доступа: <http://www.iprbookshop.ru/22431>.
2. Чернецова Е.А. Лабораторный практикум "Введение в MATLAB" [Электронный ресурс]. — Санкт-Петербург: Российский государственный гидрометеорологический университет, 2013. — 88 с. — Режим доступа: <http://www.iprbookshop.ru/12493>
3. Курячий Г.В., Маслинский К.А. Операционная система Linux. Курс лекций [Электронный ресурс]. Москва: Издательство Интернет-университета информационных технологий, 2005. — 392 с. — Режим доступа: <http://www.iprbookshop.ru/22419>.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. GNU Octave. URL: <http://www.gnu.org/software/octave/>
2. Linux. URL: <http://www.linux.org/>
3. Руководство по интерфейсу пользователя Microsoft Office Access 2010. URL: <http://office.microsoft.com/ru-ru/access-help/HA010341735.aspx>

4. Introduction to Databases | Stanford Online. URL: https://class.stanford.edu/courses/Engineering/db/2014_1/about

Перечень дополнительных информационно-методических материалов

1. Бендел Д., Нейпир Р. Использование Linux. 6-е издание. М.: Вильямс, 2003.
2. Дейт К. Дж. Введение в системы баз данных, 8-е изд. М.: Издательский дом «Вильямс», 2005.
3. Илюшечкин В.М. Основы использования и проектирования баз данных. М.: Издательство Юрайт, 2011.
4. Одиночкина С.В. Разработка баз данных в Microsoft Access 2010. СПб.: НИУ ИТМО, 2012.
5. Дейт К. Дж. SQL и реляционная теория. Как грамотно писать код на SQL. СПб.: Символ-Плюс, 2010.
6. Мишра С., Бьюли А. Секреты Oracle SQL. СПб.: Символ-Плюс, 2006.
7. Сурядный А.С. Microsoft Access 2010. Лучший самоучитель. М.: Астрель, 2012.
8. Кренке Д. Теория и практика построения баз данных, 8-е изд. СПб.: Питер, 2003.

Перечень информационных технологий и программного обеспечения

При осуществлении образовательного процесса по дисциплине в разные моменты времени используется система управления базами данных Microsoft Access одной из последних версий или СУБД MySQL, пакет численной математики Matlab (или GNU Octave), сервер под управления ОС Linux.

V. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Рекомендации по планированию и организации времени, отведенного на изучение дисциплины, описание последовательности действий обучающихся

Освоение дисциплины следует начинать с изучения рабочей учебной программы, которая содержит основные требования к знаниям, умениям и навыкам. Обязательно следует учитывать рекомендации преподавателя, данные в ходе установочных занятий. Затем – приступать к изучению отдельных разделов и тем в порядке, предусмотренном программой.

Получив представление об основном содержании раздела, темы, необходимо изучить материал с помощью рекомендуемой основной литературы. Целесообразно составить краткий конспект или схему, отображающую смысл и связи основных понятий данного раздела и включенных в него тем. Обязательно следует записывать возникшие вопросы, на которые не удалось ответить самостоятельно.

Подготовку к началу обучения включает несколько необходимых пунктов:

1) Необходимо создать для себя рациональный и эмоционально достаточный уровень мотивации к последовательному и планомерному изучению дисциплины.

2) Необходимо изучить список рекомендованной основной и дополнительной литературы и убедиться в её наличии у себя дома или в библиотеке в бумажном или электронном виде.

3) Необходимо иметь «под рукой» специальные и универсальные словари, справочники и энциклопедии, для того, чтобы постоянно уточнять значения используемых терминов и понятий. Пользование словарями и справочниками необходимо сделать привычкой. Опыт показывает, что неудовлетворительное усвоение предмета зачастую коренится в неточном, смутном или неправильном понимании и употреблении понятийного аппарата учебной дисциплины.

4) Желательно в самом начале периода обучения возможно тщательнее спланировать время, отводимое на работу с источниками и литературой по дисциплине, представить этот план в наглядной форме (график работы с датами) и в дальнейшем его придерживаться, не допуская срывов графика индивидуальной работы и «аврала» в предсессионный период. Пренебрежение этим пунктом приводит к переутомлению и резкому снижению качества усвоения учебного материала.

Рекомендации по работе с литературой

1) Всю учебную литературу желательно изучать «под конспект». Чтение литературы, не сопровождаемое конспектированием, даже пусть самым кратким – бесполезная работа. Цель написания конспекта по дисциплине – сформировать навыки по поиску, отбору, анализу и формулированию учебного материала. Эти навыки обязательны для любого специалиста с высшим образованием независимо от выбранной специальности.

2) Написание конспекта должно быть творческим – нужно не переписывать текст из источников, но пытаться кратко излагать своими словами содержание ответа, при этом максимально структурируя конспект, используя символы и условные обозначения. Копирование и «заучивание» неосмысленного текста трудоемко и по большому счету не имеет большой познавательной и практической ценности.

3) При написании конспекта используется тетрадь, поля в которой обязательны. Страницы нумеруются, каждый новый вопрос начинается с нового листа, для каждого экзаменационного вопроса отводится 1-2 страницы конспекта. На полях размещается вся вспомогательная информация – ссылки, вопросы, условные обозначения и т.д.

4) В итоге данной работы «идеальным» является полный конспект по программе дисциплины, с выделенными определениями, узловыми пунктами, примерами, неясными моментами, проставленными на полях вопросами.

5) При работе над конспектом обязательно выявляются и отмечаются трудные для самостоятельного изучения вопросы, с которыми уместно обратиться к преподавателю при посещении установочных лекций и консультаций, либо в индивидуальном порядке.

6) При чтении учебной и научной литературы всегда следить за точным и полным пониманием значения терминов и содержания понятий, используемых в тексте. Всегда следует уточнять значения по словарям или энциклопедиям, при необходимости записывать.

7) При написании учебного конспекта обязательно указывать все прорабатываемые источники, автор, название, дата и место издания, с указанием использованных страниц.

Подготовка к промежуточной аттестации по дисциплине

К аттестации допускаются студенты, которые систематически в течение всего семестра посещали и работали на занятиях и показали уверенные знания в ходе выполнении практических заданий и лабораторных работ.

Непосредственная подготовка к аттестации осуществляется по вопросам, представленным в рабочей учебной программе. Тщательно изучите формулировку каждого вопроса, вникните в его суть, составьте план ответа. Обычно план включает в себя:

— определение сущности рассматриваемого вопроса, основных положений, утверждений, определение необходимости их доказательства;

— запись обозначений, формул, необходимых для полного раскрытия вопроса;

— графический материал (таблицы, рисунки, графики), необходимые для раскрытия сущности вопроса;

— роль и значение рассматриваемого материала для практической деятельности, примеры использования в практической деятельности.

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для осуществления образовательного процесса по дисциплине необходим компьютерный класс с персональными компьютерами с доступом в сеть «Интернет». В компьютерном классе должно быть оборудование мультимедийного типа (мультимедийный проектор, настенный экран) и пластиковая доска.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Дальневосточный федеральный университет»
(ДФУ)

ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА

**УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ
РАБОТЫ ОБУЧАЮЩИХСЯ**
по дисциплине «Аналитика больших данных»
Направление подготовки 38.04.01 Экономика
магистерская программа «Государственное и муниципальное управление:
социально-экономическое развитие регионов»
Форма подготовки очная

Владивосток
2018

Характеристика заданий для самостоятельной работы обучающихся и методические рекомендации по их выполнению

1. Провести проектирование базы данных «Аэропорт». Продумать схему данных (таблиц должно быть не менее 5), связи между ними. Создать данную базу данных в СУБД Microsoft Access или в MySQL. В приложении должно быть не менее 3 форм, 2 отчетов, 1 запроса и одна главная кнопочная форма для пользователя (для MS Access).

2. Приведено описание реляционной базы данных. Требуется написать формулу на формальном языке реляционной алгебры для извлечения определенных данных из базы.

3. Приведено описание реляционной базы данных. Требуется написать запрос на языке SQL для извлечения определенных данных из базы.

4. Провести нормализацию предложенной схемы данных. Выписать все функциональные зависимости.

1.2. Каковы преимущества использования системы баз данных? Каковы недостатки использования системы баз данных?

1.3. Как вы понимаете термин *реляционная система*? Укажите различия между реляционной и нереляционной системами.

1.4. Как вы понимаете термин *модель данных*? Объясните различие между моделью данных и ее реализацией. Почему так важно это различие?

1.5. Приведите результат выполнения следующих операторов SQL выборки информации из базы данных винного погреба, представленной в табл. 11.

a)

```
SELECT WINE, PRODUCER
FROM CELLAR
WHERE BIN# = 72 ;
```

б)

```
SELECT WINE, PRODUCER
FROM CELLAR
WHERE YEAR > 2 000 ;
```

в)

```
SELECT BIN#, WINE, YEAR
FROM CELLAR WHERE READY <
2003 ;
```

г)

```
SELECT WINE, BIN#, YEAR FROM
CELLAR WHERE PRODUCER = 'Robt.
Mondavi' AND BOTTLES > 6 ;
```

1.6. Дайте собственную словесную интерпретацию типичной строки в одном из ответов на упр. 1.5, представив ее в виде истинного высказывания.

1.7. Приведите результат выполнения следующих операторов SQL внесения изменений в базу данных винного погреба, представленную в табл. 11.

a)

```
INSERT
INTO CELLAR (BIN#, WINE, PRODUCER, YEAR, BOTTLES, READY )f?
VALUES (80, 'Syrah', 'Meridian', 1998, 12, 2003 );
```

б)

```
DELETE
FROM CELLAR
WHERE READY > 2004 ;
```

в)

```
UPDATE CELLAR
SET BOTTLES = 5 WHERE
BIN# = 50 ;"
```

г)

```
UPDATE CELLAR
SET BOTTLES = BOTTLES + 2
WHERE BIN# = 50 ;
```

1.8. Напишите оператор SQL для выполнения приведенных ниже операций в базе данных винного погреба.

а) Выбрать номер ячейки, наименование вина и количество бутылок для всех вин производства Geyser Peak.

б) Выбрать номер ячейки и наименование вина для всех вин, запас которых составляет больше пяти бутылок.

в) Выбрать номер ячейки для всех красных вин.

г) Добавить три бутылки в ячейку с номером 3 0.

4.2. Даны таблицы PART и PART_STRUCTURE, определенные следующим образом.

```
CREATE TABLE PART
  ( P# P#, DESCRIPTION CHAR(100),
    PRIMARY KEY ( P# ) )
; CREATE TABLE
PART_STRUCTURE
  ( MAJOR P# P#, MINOR P# P#, QTY QTY,
    PRIMARY KEY ( MAJOR P#, MINOR P# ),
    FOREIGN KEY ( MAJOR P# ) REFERENCES
PART, FOREIGN KEY ( MINOR P# )
REFERENCES PART ) ;
```

В таблице PART_STRUCTURE показано, какие детали (MAJOR_P#) содержат другие детали (MINOR_P#) как компоненты первого уровня. Напишите на языке SQL программу для получения списка всех компонентов данной детали на все имеющихся уровнях (задача разувязки деталей). *Примечание.* Значения показанные в качестве примера на рис. 4.6, могут помочь вам более наглядно представить предложенную выше задачу. Следует отметить, что таблица PART_STRUCTURE демонстрирует, как информация о *составе изделий* (см. глав 1, раздел 1.3, подраздел "Сущности и связи") обычно представляется реляционных системах.

MAJOR_P#	MINOR_P#	QTY
P1	P2	2
P1	P3	4
P2	P3	1
P2	P4	3
P3	P5	9
P4	P5	8
P5	P6	3

Рис. 4.6. Таблица PART_STRUCTURE (значения приведены в качестве примера)

Упражнение 4.1. «Дубликаты нужны в базах данных, потому что естественным образом встречаются в реальном мире. Например, все копейки – дубликаты друг друга». Как бы вы возразили против такого аргумента?

Упражнение 4.2. Пусть r – некое отношение, и пусть bx и by – булевы выражения. Тогда существует правило (оно используется в реляционных системах для оптимизации), согласно которому $(r \text{ WHERE } bx) \cup (r \text{ WHERE } by) \equiv r \text{ WHERE } bx \text{ OR } by$ (где символом \equiv обозначается эквиваленция). Применимо ли это правило, если r – не отношение, а таблица, содержащая дубликаты?

Требования к представлению и оформлению результатов самостоятельной работы

Самостоятельная работа включает в себя повторение теоретического и практического материала дисциплины, заслушиваемого и конспектируемого в ходе аудиторных занятий; изучение основной и дополнительной литературы, указанной в рабочей учебной программе дисциплины, самоконтроль ответов на основные проблемные вопросы по темам занятий; самостоятельный повтор действий, осуществляемых в ходе выполнения лабораторных работ, в том числе при работе со специальным программным обеспечением.

Результаты самостоятельной работы представляются и оформляются в виде ответов на основные положения теоретического и практического

материала дисциплины по темам; работающего исходного кода результата решения практических заданий и задач; собственных действий, осуществляемых в ходе выполнения лабораторных работ.

Критерии оценки выполнения самостоятельной работы

Общие критерии оценки выполнения самостоятельной работы – правильность ответов на вопросы по темам теоретической части дисциплины, верность получаемых ответов в ходе решения практических заданий и задач, достижение правильного результата при осуществлении собственных действий по лабораторным работам.

Оценивание знаний в форме собеседования проводится по критериям:

- логичность изложения, знание и понимание основных теоретических и практических положений и концепций по теме;
- владение теоретическими и практическими приёмами написания, отладки и тестирования программ на языке программирования Octave, AWK
- умение написать нужный запрос на структурированном языке запросов SQL.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Дальневосточный федеральный университет»
(ДФУ)

ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Аналитика больших данных»
Направление подготовки 38.04.01 Экономика
магистерская программа «Государственное и муниципальное управление:
социально-экономическое развитие регионов»
Форма подготовки очная

Владивосток
2018

**Паспорт
фонда оценочных средств
по дисциплине «Аналитика больших данных»**

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-11: способность анализировать и использовать различные источники информации для проведения экономических расчетов	Знает	современные технологии анализа данных
	Умеет	использовать программные продукты, в т.ч. СУБД, для анализа экономических данных
	Владеет	эффективными методами разработки СУБД с помощью современных технологий программирования

№ п/п	Контролируемые разделы дисциплины	Коды и этапы формирования компетенций	Оценочные средства - наименование		
			текущий контроль	промежуточная аттестация	
1	Системы управления базами данных	ПК-11	Знает	Собеседование (УО-1)	Зачёт, вопросы 1-16
			Умеет	Проект (ПР-9)	Зачёт, проект 6-8
			Владеет	Проект (ПР-9)	Зачёт, проект 6-8

Зачётно-экзаменационные материалы

Вопросы для подготовки к зачёту по дисциплине «Аналитика больших данных»

1. Понятие базы данных. Достоинства / недостатки БД. Различные модели данных.
2. Архитектура БД. Модель ANSI / SPARC. Клиент / сервер.
3. Реляционная модель данных. Первичный ключ. Внешний ключ. Примеры.
4. Нормализация. Определения нормальных форм.
5. Функциональные зависимости и их свойства.
6. Формальный язык реляционной алгебры. Проекция, выборка (select).
Примеры.
7. Операция соединения (natural join, theta join). Декартово произведение (cross product). Примеры.
8. Операции объединения, разности, пересечения, переименования в реляционной алгебре. Примеры каждой операции.
9. Язык SQL. Все виды операций.
10. Подзапросы в SQL (после WHERE и после FROM). Примеры.
11. Агрегация данных в SQL. Условие HAVING.
12. Внутренние и внешние соединения таблиц в SQL.
13. Представления. Примеры. Триггеры.
14. Транзакции. Индексы.
15. NoSQL. MongoDB.
16. XML как модель данных, альтернативная реляционной модели. DTD, ID, IDref. Примеры.

Комплекты оценочных средств для текущей аттестации

Вопросы для собеседования по дисциплине «Аналитика больших данных»

1. Понятие базы данных. Достоинства / недостатки БД. Различные модели данных.
2. Архитектура БД. Модель ANSI / SPARC. Клиент / сервер.
3. Реляционная модель данных. Первичный ключ. Внешний ключ. Примеры.
4. Нормализация. Определения нормальных форм.
5. Функциональные зависимости и их свойства.
6. Формальный язык реляционной алгебры. Проекция, выборка (select).
Примеры.
7. Операция соединения (natural join, theta join). Декартово произведение (cross product). Примеры.
8. Операции объединения, разности, пересечения, переименования в реляционной алгебре. Примеры каждой операции.
9. Язык SQL. Все виды операций.

10. Подзапросы в SQL (после WHERE и после FROM). Примеры.
11. Агрегация данных в SQL. Условие HAVING.
12. Внутренние и внешние соединения таблиц в SQL.
13. Представления. Примеры. Триггеры.
14. Транзакции. Индексы.
15. NoSQL. MongoDB.
16. XML как модель данных, альтернативная реляционной модели. DTD, ID, IDref. Примеры.

Критерии оценки:

✓ 100-86 баллов - если ответ показывает глубокое и систематическое знание всего программного материала и структуры конкретного вопроса, а также основного содержания и новаций лекционного курса по сравнению с учебной литературой. Студент демонстрирует отчетливое и свободное владение концептуально-понятийным аппаратом, научным языком и терминологией соответствующей научной области. Знание основной литературы и знакомство с дополнительно рекомендованной литературой. Логически корректное и убедительное изложение ответа.

✓ 85-76 - баллов - знание узловых проблем программы и основного содержания лекционного курса; умение пользоваться концептуально-понятийным аппаратом в процессе анализа основных проблем в рамках данной темы; знание важнейших работ из списка рекомендованной литературы. В целом логически корректное, но не всегда точное и аргументированное изложение ответа.

✓ 75-61 - балл – фрагментарные, поверхностные знания важнейших разделов программы и содержания лекционного курса; затруднения с использованием научно-понятийного аппарата и терминологии учебной дисциплины; неполное знакомство с рекомендованной литературой; частичные затруднения с выполнением предусмотренных программой заданий; стремление логически определенно и последовательно изложить ответ.

✓ 60-50 баллов – незнание, либо отрывочное представление о данной проблеме в рамках учебно-программного материала; неумение использовать понятийный аппарат; отсутствие логической связи в ответе.

Темы проектов

по дисциплине «Аналитика больших данных»

1. Проектирование собственной базы данных в СУБД.
2. Запросы на формальном языке реляционной алгебры.
3. Язык запросов SQL.

Критерии оценки:

✓ 100-86 баллов выставляется, если студент/группа точно определили содержание и составляющие части задания, умеют аргументированно отвечать

на вопросы, связанные с заданием. Продемонстрировано знание и владение навыками самостоятельной исследовательской работы по теме. Фактических ошибок, связанных с пониманием проблемы, нет.

✓ 85-76 - баллов - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

✓ 75-61 балл – проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

✓ 60-50 баллов - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы

Описание показателей и критериев оценивания компетенций, шкал оценивания

Критерии оценки собеседования

✓ 100-86 баллов - если ответ показывает глубокое и систематическое знание всего программного материала и структуры конкретного вопроса, а также основного содержания и новаций курса по сравнению с учебной литературой. Студент демонстрирует отчетливое и свободное владение концептуально-понятийным аппаратом, научным языком и терминологией соответствующей научной области. Знание основной литературы и знакомство с дополнительно рекомендованной литературой. Логически корректное и убедительное изложение ответа.

✓ 85-76 - баллов - знание узловых проблем программы и основного содержания курса; умение пользоваться концептуально-понятийным аппаратом в процессе анализа основных проблем в рамках данной темы; знание важнейших работ из списка рекомендованной литературы. В целом логически корректное, но не всегда точное и аргументированное изложение ответа.

✓ 75-61 - балл – фрагментарные, поверхностные знания важнейших разделов программы и содержания курса; затруднения с использованием научно-понятийного аппарата и терминологии учебной дисциплины; неполное знакомство с рекомендованной литературой; частичные затруднения с выполнением предусмотренных программой заданий; стремление логически определенно и последовательно изложить ответ.

✓ 60-50 баллов – незнание, либо отрывочное представление о данной проблеме в рамках учебно-программного материала; неумение использовать понятийный аппарат; отсутствие логической связи в ответе.

Итоговая шкала оценивания

Менее 60 баллов	не зачтено	неудовлетворительно
От 61 до 75 баллов	зачтено	удовлетворительно
От 76 до 85 баллов	зачтено	хорошо
От 86 до 100 баллов	зачтено	отлично

Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины

Текущая аттестация студентов. Текущая аттестация студентов по дисциплине «Аналитика больших данных» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация по дисциплине «Аналитика больших данных» проводится в форме собеседования и выполнения проектов и осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- степень усвоения теоретических знаний - оценивается в форме собеседования;
- уровень овладения практическими умениями и навыками – оценивается в форме выполнения проектов и лабораторных работ.

Промежуточная аттестация студентов. Промежуточная аттестация студентов по дисциплине «Аналитика больших данных» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

По дисциплине предусмотрены зачёты в 3 и 4 семестрах, которые проводятся в устной форме.

Критерии выставления оценки студенту на зачёте по дисциплине «Аналитика больших данных»

Баллы (рейтинговой оценки)	Оценка зачёта (стандартная)	Требования к сформированным компетенциям
----------------------------------	--------------------------------	---

	<i>«зачтено»/ «отлично»</i>	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, правильно обосновывает принятое решение, владеет разносторонними навыками и приёмами выполнения лабораторных работ; знает современные технологии программирования; умеет разрабатывать базы данных с помощью современных технологий программирования; владеет эффективными методами разработки программных продуктов с помощью современных технологий программирования.
	<i>«зачтено»/ «хорошо»</i>	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения; знает современные технологии программирования; умеет разрабатывать программные продукты с помощью современных технологий программирования; владеет эффективными методами разработки программных продуктов с помощью современных технологий программирования.
	<i>зачтено»/ «удовлетворительно»</i>	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении лабораторных работ; частично владеет навыками разработки программ на современных языках программирования.
	<i>«не зачтено»/ «неудовлетворительно»</i>	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет лабораторные работы, практически не владеет навыками разработки программ на современных языках программирования. Как

		правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.
--	--	--