

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



подписано электронно-цифровой подписью

Программа дисциплины
Хранилища данных Б1.В.ДВ.21

Направление подготовки: 01.03.02 - Прикладная математика и информатика

Профиль подготовки: Системное программирование

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Мухтарова Т.М. , Пинягина О.В.

Рецензент(ы):

Миссаров М.Д.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Миссаров М. Д.

Протокол заседания кафедры № ____ от "____" ____ 201____ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК № ____ от "____" ____ 201____ г

Регистрационный № 9919

Казань
2019

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) старший преподаватель, б/с Мухтарова Т.М. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Tatyana.Moukhtarova@kpfu.ru ; доцент, к.н. (доцент) Пинягина О.В. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Olga.Piniaguina@kpfu.ru

1. Цели освоения дисциплины

Целью курса "Хранилища данных" является изучение основ современных методов и средств работы с хранилищами данных, проектирование структур данных, администрирования хранилищ данных и изучение технологий формирования базовых отчетов и рекомендаций руководителям предприятий по развитию информационной системы предметной области ХД.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел "Б1.В.ДВ.21 Дисциплины (модули)" основной образовательной программы 01.03.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 3 курсе, 6 семестр.

Преподавателю необходимо опираться на знания студентов, полученные при изучении предшествующих дисциплин: информатики, языков программирования и методов трансляции, математической логики и баз данных.

В результате освоения дисциплины 'Хранилища данных' студенты смогут применить полученные теоретические и практические знания при прохождении научно-исследовательской практики и подготовке отчета по практике, а также при написании магистерской диссертации.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-1 (профессиональные компетенции)	способностью использовать базовые знания естественных наук, математики и информатики, основные факты, концепции, принципы теорий, связанных с прикладной математикой и информатикой;
ОПК-2 (профессиональные компетенции)	способностью приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии;
ОПК-3 (профессиональные компетенции)	способностью к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям
ОПК-4 (профессиональные компетенции)	способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-2 (профессиональные компетенции)	способностью понимать, совершенствовать и применять современный математический аппарат
ПК-10 (профессиональные компетенции)	способностью к реализации решений, направленных на поддержку социально-значимых проектов, на повышение информационной грамотности населения, обеспечения общедоступности информационных услуг
ПК-12 (профессиональные компетенции)	способностью к планированию и осуществлению педагогической деятельности с учетом специфики предметной области в общеобразовательных и профессиональных образовательных организациях
ПК-5 (профессиональные компетенции)	способностью осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в информационно-телекоммуникационной сети "Интернет" (далее - сеть "Интернет") и в других источниках
ПК-7 (профессиональные компетенции)	способностью к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения

В результате освоения дисциплины студент:

1. должен знать:

- теоретические основы построения и использования хранилищ данных;
- принципы формирования, хранения и использования данных для выполнения задач администрирования работы предприятия;

2. должен уметь:

- использовать программные средства для построения современных хранилищ данных, обеспечивающих нужды предприятий и эффективного планирования бизнеса;
- извлекать информацию из хранилищ данных, выполнять многомерные запросы к данным, делать бизнес-прогнозы по развитию отдельных направления деятельности предприятий;

3. должен владеть:

- современным программным обеспечением, используемым для организации и управления хранилищами данных;
- инструментарием построения и организации хранилищ данных.

4. должен демонстрировать способность и готовность:

применять полученные знания в практической деятельности при написании курсовых и дипломных работ, а также магистерских диссертаций.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 6 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);
55-70 баллов - "удовлетворительно" (удов.);
54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Системы поддержки принятия решений	6	1-3	0	0	6	Письменное домашнее задание
2.	Тема 2. Концепция хранилищ данных	6	4-6	0	0	6	Письменное домашнее задание
3.	Тема 3. OLAP-кубы (On-line Analytical Processing)	6	7-9	0	0	6	Контрольная работа
4.	Тема 4. ETL (Extract - Transform - Load)	6	10-12	0	0	6	Письменное домашнее задание
5.	Тема 5. Работа с Deductor Studio (кубы, преобразования, визуализаторы)	6	13-15	0	0	6	Письменное домашнее задание
6.	Тема 6. Основы Data Mining	6	16-18	0	0	6	Контрольная работа
.	Тема . Итоговая форма контроля	6		0	0	0	Зачет
	Итого			0	0	36	

4.2 Содержание дисциплины

Тема 1. Системы поддержки принятия решений лабораторная работа (6 часа(ов)):

Обобщенная архитектура систем поддержки принятия решений (СППР). Основные задачи СППР. З класса задач анализа данных. Требования, предъявляемые к системам OLTP и СППР. Разработка моделей "звездочка" или "снежинка" для хранилища данных. Создание базы данных. Создание ER-диаграммы.

Тема 2. Концепция хранилищ данных лабораторная работа (6 часа(ов)):

Определение хранилища данных (ХД). Структура СППР с физическим ХД. Проблемы создания физического ХД. Структура СППР с виртуальным ХД. Витрины данных. Организация ХД. Оптимизация ХД. Избыточность и денормализация. Генерация данных для заполнения хранилища данными. Создание хранимых процедур

Тема 3. OLAP-кубы (On-line Analytical Processing)

лабораторная работа (6 часа(ов)):

Многомерная модель данных. Факты и измерения. Операции над OLAP-кубами: срез, вращение, агрегирование и детализация. Двенадцать правил Кодда для концептуального многомерного представления. Тест FASMI. Архитектура OLAP-систем. MOLAP, ROLAP, HOLAP. Выполнение запросов к хранилищу данных. Срезы OLAP-кубов.

Тема 4. ETL (Extract - Transform - Load)

лабораторная работа (6 часа(ов)):

Процедуры извлечения, преобразования и загрузки данных Трансформация данных. Очистка данных. Выгрузка данных в текстовые файлы и загрузка в Deductor Studio.

Тема 5. Работа с Deductor Studio (кубы, преобразования, визуализаторы)

лабораторная работа (6 часа(ов)):

Работа в программе Deductor Studio. Создание хранилища данных. Работа с OLAP-кубами. Создание сценариев в Deductor Studio. Преобразования и визуализаторы: кубы, диаграммы, детализация.

Тема 6. Основы Data Mining

лабораторная работа (6 часа(ов)):

Основные понятия Data Mining. Задачи Data Mining. Методы Data Mining. Алгоритмы Data Mining в Deductor Studio.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Системы поддержки принятия решений	6	1-3	подготовка домашнего задания	3	Письменное домашнее задание
2.	Тема 2. Концепция хранилищ данных	6	4-6	подготовка домашнего задания	3	Письменное домашнее задание
3.	Тема 3. OLAP-кубы (On-line Analytical Processing)	6	7-9	подготовка к контрольной работе	3	Контрольная работа
4.	Тема 4. ETL (Extract - Transform - Load)	6	10-12	подготовка домашнего задания	3	Письменное домашнее задание
5.	Тема 5. Работа с Deductor Studio (кубы, преобразования, визуализаторы)	6	13-15	подготовка домашнего задания	3	Письменное домашнее задание
6.	Тема 6. Основы Data Mining	6	16-18	подготовка к контрольной работе	3	Контрольная работа
Итого					18	

5. Образовательные технологии, включая интерактивные формы обучения

Д процессе изучения дисциплины 'Хранилища данных' студенты выполняют лабораторные работы по проектированию и созданию ХД, выполнению многомерных запросов к хранилищам данных в инструментальной среде SQL Express, разбирают практические ситуации, связанные с проектированием хранилищ данных, извлечению информации.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Системы поддержки принятия решений

Письменное домашнее задание , примерные вопросы:

Задание 1. Выберите предметную область, подходящую для разработки хранилища данных. Примерная схема: оперативная информация хранится в слабой сущности, плюс 3 или более сильных сущности (модель 'звезда'), либо ещё плюс справочники (модель 'снежинка') (5 баллов). Создайте таблицы в SQL Server. После создания БД проведите 'reverse engineering' - создайте диаграмму базы данных с помощью автоматизированных средств SQL Server (2 балла). ER-модель сдавать не обязательно, достаточно диаграммы из SQL Server. (Итого 7 баллов)

Тема 2. Концепция хранилищ данных

Письменное домашнее задание , примерные вопросы:

Задание 2 . Следует сгенерировать не менее 1000 записей для оперативных данных и по 10-100 для остальных таблиц. Можно использовать хранимые процедуры (если основные таблицы заполнены с помощью хранимых процедур, то 5 баллов; если при этом данные генерируются с учетом вероятностей (частот), то еще плюс 5 баллов). Можно сгенерировать данные с помощью внешней программы (написанной, например, на C#) и записать их в текстовый файл, а затем загрузить в базу (если таблица с оперативной информацией загружена из файла, то еще плюс 5 баллов). В самом крайнем случае, можно ограничиться максимум сотней строк для каждой таблицы и все данные занести вручную (1 балл). (Итого 15 баллов)

Тема 3. OLAP-кубы (On-line Analytical Processing)

Контрольная работа , примерные вопросы:

Задание 3. Напишите запросы, реализующие выборку из нескольких таблиц ? с записью полученных данных в новую таблицу и с созданием представления (3 балла). Напишите не менее 5 разнообразных запросов для получения срезов куба, с применением группировок и агрегирующих функций (5 баллов). Напишите хранимую процедуру для получения отдельной ячейки куба (3 балла). Создайте кросс-таблицу (4 балла). (Итого 15 баллов)

Тема 4. ETL (Extract - Transform - Load)

Письменное домашнее задание , примерные вопросы:

Задание 4. Произведите очистку данных: напишите запросы, которые находят пустые значения, неверные или неполные данные и, при необходимости, корректируют их (4 балла). Напишите запросы, которые преобразуют данные (4 балла). (Итого 8 баллов) Задание 5. Выгрузите подготовленные данные в один или несколько текстовых файлов. Задайте в этих файлах заголовки столбцов. Не забудьте сохранить файлы в формате ANSI. В тех файлах, где используется тип дата+время, избавьтесь от миллисекунд. (Итого 5 баллов).

Тема 5. Работа с Deductor Studio (кубы, преобразования, визуализаторы)

Письменное домашнее задание , примерные вопросы:

Задание 6. Загрузите данные в Deductor Studio из текстовых файлов, с созданием хранилища данных (15 баллов), либо без него (5 баллов). Примените к вашим данным преобразования и визуализаторы. Используйте, как минимум, визуализаторы: Куб, Диаграмма, Детализация и преобразования: Калькулятор, Преобразование даты, Фильтр, Группировка. (15 баллов).

Тема 6. Основы Data Mining

Контрольная работа , примерные вопросы:

Задание 7. Примените к вашему хранилищу методы Data mining - выделение тренда и сезонной компоненты, прогноз, ABC-анализ (10 баллов).

Итоговая форма контроля

экзамен (в 6 семестре)

Примерные вопросы к экзамену:

В процессе изучения дисциплины "Хранилища данных" студенты выполняют индивидуальные проекты по созданию хранилищ данных, включающие моделирование предметной области и программирование в SQL server, а также работу в Deductor Studio. При этом работа выполняется в основном самостоятельно. На лабораторных занятиях студенты задают вопросы по проекту, обсуждают с преподавателем проблемы, а также сдают этапы проекта.

При использовании проектного подхода проведение контрольных работ не предусмотрено. Контрольными точками можно считать этапы 3 и 7 проекта.

Вопросы к экзамену:

1. Какие задачи решают СППР?
2. На какие типы делятся задачи анализа данных?
3. Из чего состоит архитектура СППР?
4. Каковы требования к системам OLTP и СППР?
5. Что такое хранилище данных?
6. Что такое виртуальное и физическое хранилище данных?
7. Что такое витрина данных?
8. Из чего состоит хранилище данных?
9. Какие потоки данных имеются в хранилище данных?
10. Какие есть приемы оптимизации хранилищ данных?
11. Для чего нужна денормализация данных?
12. Что такое нисходящая денормализация, восходящая денормализация, внутритабличная денормализация?
13. Какие операции можно производить над гиперкубом?
14. Поясните все составные части теста FASMI.
15. Приведите плюсы и минусы MOLAP
16. Приведите плюсы и минусы ROLAP
17. Основные цели и задачи процесса ETL
18. Извлечение данных в ETL
19. Выбор используемых источников данных в ETL. Особенности организации процесса извлечения данных
20. Очистка данных в ETL
21. Основные виды проблем в данных, из-за которых они нуждаются в очистке
22. Преобразование данных в ETL
23. Преобразование структуры данных
24. Агрегирование данных
25. Перевод значений
26. Создание новых данных
27. Преимущества и недостатки отказа от хранилищ данных
28. Квантование (дискретизация)
29. Нормализация. Десятичное масштабирование. Минимаксная нормализация.
30. Нормализация с помощью стандартного отклонения. Нормализация с помощью поэлементных преобразований
31. Что такое Data Mining?
32. Основные задачи Data Mining
33. Описательные и предсказательные задачи
34. Supervised learning и unsupervised learning
35. Этапы интеллектуального анализа данных
36. Методы интеллектуального анализа данных

7.1. Основная литература:

- 1.Барсегян А. А. Анализ данных и процессов: Учебное пособие / Барсегян А.А., Куприянов М.С., Холод И.И. - СПб:БХВ-Петербург, 2009. - 512 с. ISBN 978-5-9775-0368-6
<http://znanium.com/catalog.php?bookinfo=350638>
2. Бондарь А. Г. Microsoft SQL Server 2012: Пособие / Бондарь А.Г. - СПб:БХВ-Петербург, 2013. - 608 с. ISBN 978-5-9775-0501-7
<http://znanium.com/bookread2.php?book=941206>
- 3.Базы данных. В 2-х кн. Кн. 2. Распределенные и удаленные базы данных: Учебник / В.П. Агальцов. - М.: ИД ФОРУМ: НИЦ Инфра-М, 2013. - 272 с.: ил.; 60x90 1/16. - (Высшее образование). (переплет) ISBN 978-5-8199-0394-0
<http://znanium.com/bookread2.php?book=372740>
- 4.Култыгин, О. П. Администрирование баз данных. СУБД MS SQL Server [Электронный ресурс] : учеб. пособие / О. П. Култыгин. - М.: МФПА, 2012. - 232 с. - (Университетская серия). - ISBN 978-5-4257-0026-1.
<http://znanium.com/bookread2.php?book=451114>
- 5.Ишмухаметов, Ш.Т., Рубцова, Р.Г. Лабораторные работы по курсу 'Хранилища данных', 2012. - URL: http://kpfu.ru/docs/F855838989/Data_Bases_Labs_2012.pdf
- 6.Информационные системы в экономике: Учеб. пособие / Под ред. Д.В. Чистова. - М.: ИНФРА-М, 2009. - 234 с.:
<http://www.znanium.com/bookread.php?book=154831>

7.2. Дополнительная литература:

- 1.Степанов, Роман Григорьевич. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т.?Казань: Казанский государственный университет, 2009.?110 с.
- 2.Базы данных : теория и практика : учебник для бакалавров : для студентов вузов, обучающихся по направлениям 'Информатика и вычислительная техника' и 'Информационные системы' / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской .? Издание 2-е .? Москва : Юрайт, 2012 .? 463 с. : ил., табл., схемы ; 22 см. ? (Бакалавр, Базовый курс) (Учебно-методическое объединение рекомендует) .? На 4-й с. обл. авт.: Советов Б.Я., д.т.н., проф., засл. деят. науки и техники РФ, акад. Рос. акад. образования, Цехановский В.В., к.т.н., доц., Чертовской В.Д., д.т.н., проф. ? Библиогр.: с. 459-460 (49 назв.) .? ISBN 978-5-9916-2010-9 ((в пер.)) , 1000.
- 3.Пинягина, О. В. Практикум по курсу 'Базы данных': [учебное пособие] / О. В. Пинягина, И. А. Фукин; Казан. (Приволж.)федер. ун-т.?Казань: Казанский университет, 2012.?91 с.
- 4.Пирогов, В. Ю. Информационные системы и базы данных: организация и проектирование: учеб. пособие / В. Ю. Пирогов. ? СПб.: БХВ-Петербург, 2009. ? 528 с.: ил. ? (Учебная литература для вузов). - ISBN 978-5-9775-0399-0.
<http://znanium.com/bookread2.php?book=350672>
- 5.Дунаев В.В. Базы данных. Язык SQL для студента. - 2-е изд., доп. и перераб.- СПб.: БХВ-Петербург, 2007. - ISBN 978-5-9775-0113-2.
<http://znanium.com/bookread2.php?book=350372>

7.3. Интернет-ресурсы:

Deductor Academic - <http://www.basegroup.ru/download/deductor/>

Методические материалы для выполнения практических заданий, часть 1 -
http://kek.ksu.ru/EOS/DW/DW_pract.doc

Методические материалы для выполнения практических заданий, часть 2 -

http://kek.ksu.ru/EOS/DW/DW_pract2.doc

Регламент курса - http://kek.ksu.ru/EOS/DW/DW_reg.doc

Страница курса на сайте КЭК - <http://kek.ksu.ru/EOS/DW/index.html>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Хранилища данных" предполагает использование следующего материально-технического обеспечения:

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань" , доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Компьютерные классы лаборатории малой вычислительной техники Института ВМиИТ, оснащенные мультимедийным оборудованием.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.03.02 "Прикладная математика и информатика" и профилю подготовки Системное программирование .

Автор(ы):

Мухтарова Т.М. _____

Пинягина О.В. _____

"__" 201__ г.

Рецензент(ы):

Миссаров М.Д. _____

"__" 201__ г.