

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



Программа дисциплины
Хранилища данных БЗ.В.1

Направление подготовки: 080500.62 - Бизнес-информатика

Профиль подготовки:

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Пинягина О.В.

Рецензент(ы):

Миссаров М.Д.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Миссаров М. Д.

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

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

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

Регистрационный No 9107615

Казань

2015

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) 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 " Б3.В.1 Профессиональный" основной образовательной программы 080500.62 Бизнес-информатика и относится к вариативной части. Осваивается на 2 курсе, 4 семестр.

Курс лекций и практических занятий "Хранилища данных" изучается на 2 курсе в 4 семестре. При составлении лекций преподавателю необходимо учесть 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
ОК-12 (общекультурные компетенции)	осознает сущность и значение информации в развитии современного общества; владеет основными методами, способами и средствами получения, хранения, переработки информации
ОК-13 (общекультурные компетенции)	имеет навыки работы с компьютером как средством управления информацией, способен работать с информацией в глобальных компьютерных сетях;
ОК-16: (общекультурные компетенции)	способен работать с информацией из различных источников;
ПК-20 (профессиональные компетенции)	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования;
ПК-21 (профессиональные компетенции)	готовить научно-технические отчеты, презентации, научные публикации по результатам выполненных исследований.

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 направлений деятельности предприятий;

3. должен владеть:

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

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) 144 часа(ов).

Форма промежуточного контроля дисциплины зачет в 4 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	4	1-3	2	0	4	домашнее задание
2.	Тема 2. Концепция хранилищ данных	4	4-6	3	0	6	домашнее задание
3.	Тема 3. OLAP-кубы (On-line Analytical Processing)	4	7-9	3	0	6	домашнее задание
4.	Тема 4. ETL (Extract - Transform - Load)	4	10-12	4	0	6	домашнее задание

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
5.	Тема 5. Основы Data Mining	4	13-15	4	0	6	домашнее задание
6.	Тема 6. Работа с Deductor Studio (кубы, преобразования, визуализаторы)	4	16-18	2	0	8	домашнее задание
	Тема . Итоговая форма контроля	4		0	0	0	зачет
	Итого			18	0	36	

4.2 Содержание дисциплины

Тема 1. Системы поддержки принятия решений

лекционное занятие (2 часа(ов)):

Обобщенная архитектура систем поддержки принятия решений (СППР). Основные задачи СППР. 3 класса задач анализа данных. Требования, предъявляемые к системам OLTP и СППР.

лабораторная работа (4 часа(ов)):

Разработка моделей "звездочка" или "снежинка" для хранилища данных. Создание базы данных. Создание ER-диаграммы.

Тема 2. Концепция хранилищ данных

лекционное занятие (3 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.

лабораторная работа (6 часа(ов)):

Генерация данных для заполнения хранилища данными. Создание хранимых процедур.

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

лекционное занятие (3 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. Схемы "звезда" и "снежинка".

лабораторная работа (6 часа(ов)):

Выполнение запросов к хранилищу данных. Срезы OLAP-кубов.

Тема 4. ETL (Extract - Transform - Load)

лекционное занятие (4 часа(ов)):

Процедуры извлечения, преобразования и загрузки данных.

лабораторная работа (6 часа(ов)):

Трансформация данных. Очистка данных.

Тема 5. Основы Data Mining

лекционное занятие (4 часа(ов)):

Основные понятия Data Mining. Задачи Data Mining. Методы Data Mining.

лабораторная работа (6 часа(ов)):

Выгрузка данных в текстовые файлы и загрузка в Deductor Studio.

Тема 6. Работа с Deductor Studio (кубы, преобразования, визуализаторы)**лекционное занятие (2 часа(ов)):**

Работа в программе Deductor Studio с OLAP-кубами. Преобразования и визуализаторы. Алгоритмы Data Mining в Deductor Studio.

лабораторная работа (8 часа(ов)):

Создание сценариев в Deductor Studio. Визуализация: кубы, диаграммы, детализация.

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	4	1-3	подготовка домашнего задания	13	домашнее задание
2.	Тема 2. Концепция хранилищ данных	4	4-6	подготовка домашнего задания	14	домашнее задание
3.	Тема 3. OLAP-кубы (On-line Analytical Processing)	4	7-9	подготовка домашнего задания	14	домашнее задание
4.	Тема 4. ETL (Extract - Transform - Load)	4	10-12	подготовка домашнего задания	14	домашнее задание
5.	Тема 5. Основы Data Mining	4	13-15	подготовка домашнего задания	15	домашнее задание
6.	Тема 6. Работа с Deductor Studio (кубы, преобразования, визуализаторы)	4	16-18	подготовка домашнего задания	20	домашнее задание
	Итого				90	

5. Образовательные технологии, включая интерактивные формы обучения

В соответствии с требованиями ФГОС удельный вес занятий, проводимых в активных и интерактивных формах, составляет не менее 40% аудиторных занятий. Так, в процессе изучения дисциплины "Хранилища данных" студенты выполняют лабораторные работы по проектированию и созданию ХД, выполнению многомерных запросов к хранилищам данных в инструментальной среде Oracle Express, разбирают практические ситуации, связанные с проектированием хранилищ данных, извлечению информации, выступают со стендовыми докладами. До 50% лекционных и практических занятий проходят с использованием презентаций MS PowerPoint.

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**

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

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 С#) и записать их в текстовый файл, а затем загрузить в базу (если таблица с оперативной информацией загружена из файла, то ещё 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)

домашнее задание , 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)

домашнее задание , примерные вопросы:

Задание 4. Произведите очистку данных: напишите запросы, которые находят пустые значения, неверные или неполные данные и, при необходимости, корректируют их (5 баллов). Напишите запросы, которые преобразуют данные (4 балла). (Итого 9 баллов)

Тема 5. Основы Data Mining

домашнее задание , примерные вопросы:

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 6. Работа с Deductor Studio (кубы, преобразования, визуализаторы)

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). Задание 7. Примените к вашим данным преобразования и визуализаторы. Используйте, как минимум, визуализаторы: Куб, Диаграмма, Детализация и преобразования: Калькулятор, Преобразование даты, Фильтр, Группировка. (15 баллов).

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

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.

Окончательный вид проекта студенты демонстрируют и защищают на зачете.

При использовании проектного подхода проведение контрольных работ не предусмотрено.

Контрольными точками можно считать этапы 3 и 7 проекта.

Вопросы к зачету:

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. Автоматизированные информационные технологии в экономике: учебник / под ред. Г. А. Титоренко. М.: ЮНИТИ, 2004. 399 с
2. Информационные системы в экономике: Учеб. пособие / Под ред. Д.В. Чистова. - М.: ИНФРА-М, 2009. - 234 с.:
<http://www.znaniyum.com/bookread.php?book=154831>

3. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. ? 3-е изд., перераб. и доп. ? СПб.: БХВ-Петербург, 2009. ? 512 с.: ил. + CD-ROM ? (Учебная литература для вузов).

<http://www.znaniyum.com/bookread.php?book=350638>

4. 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

5. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. - 578 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-9558-0191-9 <http://www.znaniyum.com/bookread.php?book=251791>

7.2. Дополнительная литература:

1. Туманов, В. Е. Проектирование хранилищ данных для систем бизнес-аналитики: учебное пособие / В. Е. Туманов. ? Москва: Интернет-Университет Информационных Технологий: БИНОМ. Лаборатория знаний, 2011. ? 615 с.

2. Ишмухаметов, Ш.Т., Рубцова, Р.Г. Лабораторные работы по курсу "Хранилища данных", 2012. - URL: http://kpfu.ru/docs/F855838989/Data_Bases_Labs_2012.pdf

3. Кузин, Александр Владимирович. Базы данных: учеб. пособие для студентов вузов, обучающихся по направлению подгот. дипломир. специалистов 654600 "Информатика и вычислит. техника" / А. В. Кузин, С. В. Левонисова. ? Москва: Академия, 2005. ? 314 с.

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:

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

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 080500.62 "Бизнес-информатика" .

Автор(ы):

Пинягина О.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Миссаров М.Д. _____

"__" _____ 201__ г.