

Министерство образования и науки Российской Федерации
федеральное государственное бюджетное образовательное
учреждение высшего образования
«Казанский национальный исследовательский технический университет
им. А.Н. Туполева-КАИ»
(КНИТУ-КАИ)

Институт Компьютерных технологий и защиты информации

Кафедра Компьютерных систем

УТВЕРЖДАЮ

Ответственный за ОП

Верш И.С.Вершинин
"31" 08 2017 г.

Регистрационный номер: 4010 -

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

17/М - 096

для проведения промежуточной аттестации обучающихся дисциплине

«Проектирование и внедрение корпоративных хранилищ данных»

Индекс по учебному плану: Б1.В.ДВ.02.01

Направление подготовки: 09.04.01 «Информатика и вычислительная техника»

Квалификация: магистр

Магистерские программы:

Информационное и программное обеспечение автоматизированных систем

Вид профессиональной деятельности: научно-исследовательская

Заведующий кафедрой ПМИ С.С. Зайдуллин

Разработчики:

профессор кафедры ПМИ, д.т.н. Л.Ю. Емалетдинова

доцент кафедры ПМИ, к.т.н. Р.И. Габитов

Казань 2017 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения промежуточной аттестации обучающихся дисциплине

«Проектирование и внедрение корпоративных хранилищ данных»

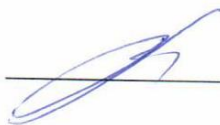
Индекс по учебному плану: Б1.В.ДВ.02.01 Направление подготовки: 09.04.01 «Информатика и вычислительная техника» Квалификация: магистр Магистерские программы: Информационное и программное обеспечение автоматизированных систем Вид профессиональной деятельности: научно-исследовательская

Содержание фонда оценочных средств соответствует требованиям федерального государственного стандарта высшего образования (ФГОС ВО) по направлению подготовки 09.04.01 «Информатика и вычислительная техника». ФОС достаточно полно отражает дисциплину и актуален, необходим для профессиональной деятельности обучающихся, установленных ФГОС ВО. В наличии имеются оценочные средства для проведения различных форм контроля, формы заданий достаточно разнообразны, имеются контекстные задания для различного уровня подготовки, количество вариантов достаточно. Приведенные в ФОС задачи аналогичны решаемым задачам на производстве. Существенных недостатков не обнаружено.

Представленные материалы соответствуют ФГОС ВО по направлению: 09.04.01 «Информатика и вычислительная техника» рекомендуется для использования в учебном процессе.

Рассмотрено на заседании учебно-методической комиссии института КТЗИ от 31 августа 2017 г., протокол №.8

Председатель УМК института КТЗИ _____ В.В. Родионов



Содержание

ВВЕДЕНИЕ	4
1. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ	5
2. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ	5
3. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ В ПРОЦЕССЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
4. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА РАЗЛИЧНЫХ ЭТАПАХ ИХ ФОРМИРОВАНИЯ, ОПИСАНИЯ ШКАЛЫ ОЦЕНИВАНИЯ	6
5. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРУ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ	9
6 КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ПРОЦЕССЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ	11
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ И ДОПОЛНЕНИЙ	25

Введение

Фонд оценочных средств промежуточной аттестации предназначен для оценки запланированных результатов по дисциплине «Проектирование и внедрение корпоративных хранилищ данных».

В соответствии с требованиями ФГОС ВПО для аттестации обучающихся на соответствие их персональных достижений поэтапным требованиям соответствующей образовательной программы разработан фонд оценочных средств промежуточной аттестации (ФОС ПА), включающий типовые задания, контрольные вопросы, тесты, позволяющие оценить уровень приобретённых студентом компетенций, знаний и умений, оценить уровень владения полученными навыками.

Фонд оценочных средств является составной частью учебных и методических документов, обеспечивающих реализацию конкретной образовательной программой.

Задачи ФОС ПА по дисциплине «Проектирование и внедрение корпоративных хранилищ данных»:

- оценка достижений студентов в процессе изучения дисциплины в соответствии с разработанными и принятыми критериями по каждому виду контроля;
- управление процессом приобретения студентами необходимых знаний, умений, навыков и формирования компетенций, определённых в ФГОС ВПО по соответствующему направлению подготовки.

Основные принципы ФОС ПА по дисциплине «Проектирование и внедрение корпоративных хранилищ данных»:

- валидность (объекты оценки соответствуют поставленным целям обучения);
- надёжность (точность, степень постоянства, стабильности, устойчивости результатов оценивания при повторных предъявлениях);
- системность оценивания (циклический характер оценивания);
- соответствие содержания материалов оценочных средств уровню и стадии обучения;
- наличие чётко сформулированных критериев оценки для каждого контрольного мероприятия;
- максимальная объективность используемых процедур и методов оценки;
- использование ФОС ПА не только в качестве средства оценивания, но и обучения.

1. Формы промежуточной аттестации по дисциплине

Дисциплина изучается в одном семестре на первом курсе магистратуры и завершается промежуточной аттестацией в 1 семестре.

2. Оценочные средства для промежуточной аттестации

Таблица 1

№ п.п.	№ раздела / модуля	Код ФОС ПА	Вид оценочных средств	Примечание	Формируемые компетенции
1	Разделы 1-4	ФОСТК-1	ТПА-1	Тест промежуточной аттестации (для экзамена)	ПК-7
2	Разделы 1-4	ФОСТК-2	ТПА-2	Тест промежуточной аттестации (для экзамена)	ПК-7
3	Разделы 1-4	ФОСТК-3	ТПА-3	Тест промежуточной аттестации (для экзамена)	ПК-7

3. Перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины

Перечень компетенций и их составляющих, которые должны быть сформированы при изучении темы соответствующего раздела дисциплины «Проектирование и внедрение корпоративных хранилищ данных», представлен в таблице 2.

Таблица 2

Перечень компетенций и этапы их формирования
в процессе освоения дисциплины

№ п/п	Этап формирования (семестр)	Наименование раздела	Код формируемой компетенции (составляющей компетенции)		Форма промежуточной аттестации
1.	1	Основные понятия и определения многомерных хранилищ данных	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-1
2.	1	Способы реализации гиперкубов на основе OLAP	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-1

3	1	Витрины данных	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-2
4	1	ETL процессы	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-2
5	1	Оптимизация работы с хранилищем данных	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-3
6	1	Технология Data Mining	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ТК-3
7	1	Экзамен	ПК-7	ПК-7.3, ПК-7.У, ПК-7.В	ФОС ПА - комплексное задание

4. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описания шкалы оценивания

Показатели и критерии оценивания сформированности компетенций на зачете, приведены в таблице 3.

Таблица 3

Показатели и критерии оценивания сформированности компетенций на зачете

№ п/п	Этап формирования (семестр)	Код формируемой компетенции (составляющей компетенции)		Критерии оценивания	Показатели оценивания (планируемые результаты обучения)		
					Пороговый уровень	Продвинутый уровень	Превосходный уровень
1.	1	ПК-7	ПК-7.З, ПК-7.У, ПК-7.В	Теоретические навыки	теоретические основы построения хранилищ данных методику проектирования хранилищ данных принципы работы с хранилищами данных структуру и состав систем подготовки принятия решений	теоретические основы построения хранилищ данных методику проектирования хранилищ данных язык запросов к многомерным хранилищам данных MDX принципы работы с хранилищами данных структуру и состав систем подготовки принятия решений технологии разработки и внедрения хранилищ данных	теоретические основы построения хранилищ данных методику проектирования хранилищ данных язык запросов к многомерным хранилищам данных MDX принципы работы с хранилищами данных структуру и состав систем подготовки принятия решений технологии разработки и внедрения хранилищ данных задачи и методы технологии Data Mining для многомерных хранилищ

2.	1	ПК-7	ПК-7.З, ПК-7.У, ПК-7.В	Практические навыки	проектировать денормализованные реляционные хранилища данных по схеме «звезда»; заполнять хранилища данными; создавать многомерные OLAP-кубы на основе технологии MOLAP;	проектировать денормализованные реляционные хранилища данных по схеме «звезда»; заполнять хранилища данными; создавать многомерные OLAP-кубы на основе технологии MOLAP; использовать MDX операторы для работы с хранилищами данных; просматривать содержимое OLAP-кубов средствами редактора кубов SQL Server; просматривать содержимое OLAP-кубов средствами Microsoft Office;	проектировать денормализованные реляционные хранилища данных по схеме «звезда»; заполнять хранилища данными; создавать многомерные OLAP-кубы на основе технологии MOLAP; использовать MDX операторы для работы с хранилищами данных; просматривать содержимое OLAP-кубов средствами редактора кубов SQL Server; просматривать содержимое OLAP-кубов средствами Microsoft Office; публиковать сводные таблицы и диаграммы на Web-страницах
----	---	------	------------------------------	---------------------	--	---	---

Формирование оценки при промежуточной аттестации по итогам освоения дисциплины зависит от уровня освоения компетенций, которые обучающийся должен освоить по данной дисциплине. Связь между итоговой оценкой и уровнем освоения компетенций (шкала оценивания) представлена в таблице 4.

Таблица 4

Описание шкалы оценивания

Шкала оценивания		Описание оценки в требованиях к уровню и объему компетенций
Словесное выражение	Выражение в баллах	
Зачтено	от 86 до 100	Освоен превосходный уровень всех компетенций (составляющих компетенций)
Зачтено	от 71 до 85	Освоен продвинутый уровень всех компетенций (составляющих компетенций)
Зачтено	от 51 до 70	Освоен пороговый уровень всех компетенций (составляющих компетенций)
Не зачтено	до 51	Не освоен пороговый уровень всех компетенций (составляющих компетенций)

5. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Формирование оценки по результатам текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины приведено в таблице 5.

Таблица 5

Формирование оценки по итогам освоения дисциплины

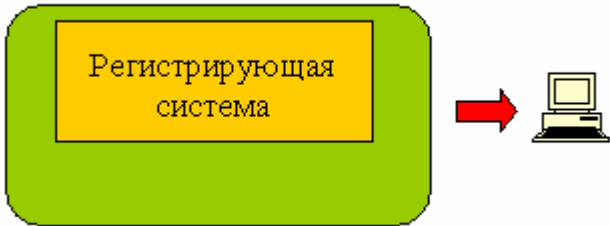
Наименование контрольного мероприятия	Рейтинговые показатели				
	I аттестация	II аттестация	III аттестация	по результатам текущего кон- троля	по итогам промежуточной аттестации (зачета /экзамена)
Раздел 1. Основные понятия и определения многомерных хранилищ данных	5			5	
Тест текущего контроля по разделу	5			5	
Раздел 2. Способы реализации гиперкубов на основе OLAP		10		10	
Тест текущего контроля по разделу		5		5	
Защита лабораторных работ		5		5	
Раздел 3. Витрины данных		10		10	
Тест текущего контроля по разделу		5		5	
Защита лабораторных работ		5		5	
Раздел 4. ETL процессы			10	10	
Тест текущего контроля по разделу			5	5	
Защита лабораторных работ			5	5	
Раздел 5. Оптимизация работы с хранилищем данных			10	10	
Тест текущего контроля по разделу			5	5	
Защита лабораторных работ			5	5	
Раздел 5. Технология Data Mining			5	5	
Тест текущего контроля по разделу			5	5	
Итоговая аттестация (экзамен):					50
– итоговый тест по дисциплине					20
– ответы на контрольные вопросы в письменной форме					30

6 Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения дисциплины

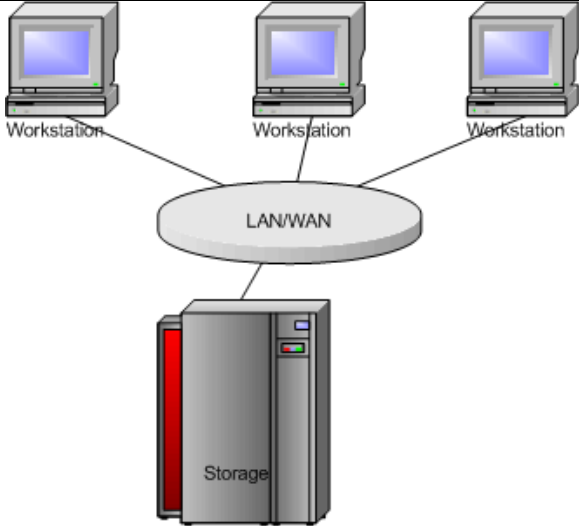
6.1. Оценочные средства для промежуточной аттестации ФОСТК-1

6.1.1. Тестовые вопросы ТПА-1

Хранилища данных и системы хранения

№	Вопросы	Ответы
1-1	Хранилище данных – это	1. территориально распределенная универсальная интегрированная база данных, предназначенная для синхронизации данных и организованная согласно схеме «снежинка»; 2. компьютерные системы, осуществляющие учет операций и собственно доступ к базам транзакций; 3. предметно-ориентированная, интегрированная, содержащая исторические данные, неразрушаемая совокупность данных, предназначенная для поддержки принятия управленческих решений.
1-2	Данная архитектура соответствует - 	1. виртуальному ХД 2. двухуровневому ХД 3. трехуровневому ХД

1-3	Какая архитектура хранилища данных подразумевает построение витрин данных (data mart) без создания центрального хранилища	1. виртуальное ХД 2. двухуровневое ХД 3. трехуровневое ХД
1-4	Что содержит третий уровень трехуровневого ХД?	1. он является источником аналитической информации для оперативного управления 2. здесь подготавливаются данные для последующей загрузки в центральное хранилище. 3. набор предметно-ориентированных витрин данных, источником информации для которых является центральное хранилище данных.
1-5	Система, подключенная напрямую к серверу, называется	1. Direct Attached Storage (DAS) 2. Network Attached Storage (NAS) 3. Storage Area Network (SAN)
1-6	Данная схема соответствует системе хранения данных -	1. Direct Attached Storage (DAS) 2. Network Attached Storage (NAS) 3. Storage Area Network

		(SAN)
1-7	Особенностями какой системы хранения являются: выделенный файловый сервер и независимость доступа к данным от ОС и платформы;	1. Direct Attached Storage (DAS) 2. Network Attached Storage (NAS) 3. Storage Area Network (SAN)
1-8	Особенностями какой системы хранения являются: независимость топологии от сторедаж-систем и серверов; удобное централизованное управление; отсутствие конфликта с трафиком LAN/WAN;	1. Direct Attached Storage (DAS) 2. Network Attached Storage (NAS) 3. Storage Area Network (SAN)

Создание многомерного ХД

№	Вопросы	Ответы
2-1	Факт - это	1. множество объектов одного или нескольких типов, организованных в виде иерархической структуры и обеспечивающих инфор-

		мационный контекст числового показателя 2. величина (обычно числового типа), которая собственно и является предметом анализа 3. атомарная структура куба, соответствующая конкретному значению некоторого показателя
2-2	Какой тип фактов моделирует состояние объекта в данный момент времени?	1. События 2. Мгновенные снимки 3. Совокупные мгновенные снимки
2-3	Технология комплексного многомерного анализа данных получила название	1. OLAP 2. OLTP 3. SQL
2-4	Какие системы хранят данные в специализированных многомерных структурах?	1. ROLAP 2. HOLAP 3. MOLAP
2-5	Какая таблица ХД содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться	1. таблица измерений 2. таблица фактов 3. таблица аргументов
2-6	Схема «звезда» должна содержать	1. одну таблицу фактов и несколько таблиц измерений 2. несколько таблиц фактов и одну таблицу измерений 3. несколько таблиц фактов и несколько таблиц измерений
2-7	Иерархии, в которых число уровней определено её структурой и неизменно, называются	1. сбалансированными 2. несбалансированными 3. неровными
2-8	Суть какого процесса состоит в вычислении значений, соответствующих меткам остальных уровней иерархии на основе фактов нижнего уровня?	1. интерпретации 2. декомпозиции 3. агрегации

6.1.2. Тестовые вопросы ТПА-2

Витрины данных и ETL-процессы.

№	Вопросы	Ответы
3-1	Витрина данных - это	1. структура, где подготавливаются данные для последующей загрузки в центральное хранилище 2. атомарная структура куба, соответствующая конкретному значению некоторого показателя 3. срез хранилища данных, представляющий собой массив тематической, узконаправленной информации
3-2	Какой метод перевычисления таблиц итогов предпочтительнее, если вы располагаете небольшой базой данных и быстрым сервером?	1. полное перевычисление 2. инкрементальное с периодическим перестроением таблиц 3. метод скользящего окна
3-3	Для чего используется специальный инструмент, называемый ETL?	1. для инкрементального перевычисления таблиц итогов с периодическим перестроением таблиц 2. для извлечения, трансформации и загрузки данных 3. для формирования и выполнения MDX-запросов к многомерному хранилищу данных
3-4	Что такое фаза процесса ETL?	1. представляет собой отдельный SQL-запрос, который выполняет единичные действия по перегрузке, преобразованию и выборке данных 2. это реализация потока данных от единственного набора данных источника до одного или нескольких

		наборов данных ХД 3. подпроцесс, обеспечивающий решение определённой задачи в рамках ETL-процесса
3-5	При выявлении ошибок в ХД данные, не прошедшие проверку	1. удаляются 2. поступают в таблицу с префиксом ster 3. поступают в таблицу с префиксом stac
3-6	Распределение данных на несколько потоков перед вставкой в ХД требуется для того,	1. чтобы разделить новые данные и записи, которые должны обновить или дополнить ранее поступившую информацию в ХД 2. разделить правильные и ошибочные данные, т.е. не удовлетворяющие существующим бизнес-правилам 3. распределить данные по удалённым серверам
3-7	Какие фазы ETL-процесса подвергаются оптимизации чаще других?	1. очистки данных, распределения данных и вставки 2. извлечения, преобразования и распределения данных 3. очистки данных, генерации суррогатных ключей и фаза вставки данных в ХД
3-8	Какие потоки оптимизирует фаза вставки данных в ХД?	1. STER и STAC 2. STIN и STUP 3. STCF и STTM

6.1.3.Тестовые вопросы ТПА-3

Оптимизация ХД.

№	Вопросы	Ответы
---	---------	--------

4-1	В чем заключается проблема производительности ХД?	1. в уменьшении скорости выполнения процессов в хранилище 2. в увеличении объема ХД 3. в наличии ошибок и несоответствий в данных
4-2	Разделение информации в процессе управления жизненным циклом (ILM) производится на-	1. достоверную и ошибочную 2. локальную и глобальную 3. оперативную и квазиоперативную
4-3	Орфографические ошибки, избыточность и дубликаты, противоречивые значения – это проблемы качества данных	1. уровня элементов множественных источников 2. уровня элементов отдельных источников 3. уровня схемы отдельных источников
4-4	Какой из перечисленных пунктов не является этапом очистки данных?	1. анализ данных 2. определение порядка и правил преобразования данных 3. слияние данных
4-5	Какие из перечисленных средств не являются средствами анализа данных?	1. средства мониторинга данных 2. средства профайлинга данных 3. средства data mining
4-6	Семантический разрыв – это:	1. непонимание информации, которая заключена в данных. 2. нарушение ограничения семантической целостности ХД 3. совокупность орфографических и синтаксических ошибок в данных

4-7	Вавилонский семантический разрыв - это	1. разрыв между параллельными потоками данных, поступающими к аналитику 2. разрыв в интерпретации имен понятий 3. разрыв в интерпретации связей между понятиями
4-8	Кроссязыковый семантический разрыв – это:	1. разрыв между параллельными потоками данных, поступающими к аналитику 2. разрыв в интерпретации имен понятий 3. разрыв в интерпретации связей между понятиями

Оптимизация ХД.

№	Вопросы	Ответы
5-1	Data Mining – это:	1. процесс формирования MDX-запросов к многомерному ХД 2. процесс обнаружения в данных ранее неизвестных, практически полезных знаний 3. добыча данных из хранилища при помощи SQL-запросов
5-2	В каких сферах не применяются технологии Data Mining?	1. научные исследования 2. бизнес-задачи 3. задачи проектирования реляционных БД
5-3	Классификация, Кластеризация и Ассоциация являются задачами:	1. Data Mining 2. ETL 3. OLAP-анализа
5-4	Не являются стадиями Data Mining:	1. свободный поиск 2. синтез исключений 3. прогностическое моделирование
5-5	Статистическими методами Data mining являются:	1. корреляционный и регрессионный анализ, факторный анализ

		2. искусственные нейронные сети, эволюционное программирование, генетические алгоритмы 3. нечеткая логика, деревья решений, системы обработки экспертных знаний
5-6	Кластеризация предназначена для:	1. упорядочения по некоторому принципу множества объектов, которые имеют сходные признаки 2. определения тенденций динамики конкретного объекта или события на основе ретроспективных данных, т.е. анализа его состояния в прошлом и настоящем. 3. разбиения совокупности объектов на однородные группы
5-7	Точность прогноза характеризуется:	1. трендом 2. горизонтом 3. ошибкой
5-8	Какие методы не относятся к представлению данных в 4+ измерениях?	1. лица Чернова 2. карты Кохонена 3. параллельные координаты

6.2. Оценочные средства для промежуточной аттестации экзаменационные билеты ФОСПА-1

6.2.Экзаменационные билеты ФОСПА-1

Билет №1

Вопрос 1	Вопрос 2
Хранилище данных. Основные понятия и определения.	Проблема очистки данных на уровне элемента для единственного источника.

Билет №2

Вопрос 1	Вопрос 2
Основные виды архитектуры хранилища данных: виртуальная, двухуровневая и трехуровневая.	Проблема очистки данных для нескольких источников.

Билет №3

Вопрос 1	Вопрос 2
Основные системы хранения данных. Direct Attached Storage (DAS).	Основные этапы очистки данных.

Билет №4

Вопрос 1	Вопрос 2
Основные системы хранения данных. Storage Area Network (SAN).	Семантический разрыв: понятия «концепт» и «денотат».

Билет №5

Вопрос 1	Вопрос 2
Понятие многомерной базы данных и гиперкуба.	Семантические разрывы и их виды.

Билет №6

Вопрос 1	Вопрос 2
Понятие измерения и иерархии измерений.	Понятие и назначение Data Mining.

Билет №7

Вопрос 1	Вопрос 2
Технологии OLTP и OLAP. Их сравнительная характеристика.	Назовите сферы применения технологии Data Mining.

Билет №8

Вопрос 1	Вопрос 2

Построение многомерных баз данных. Технология «звезда» и «снежинка».	Технология Data Mining: Web Mining, Text Mining, Call Mining.
---	--

Билет №9

Вопрос 1	Вопрос 2
Иерархия измерений и ее применение для агрегирования данных.	Основные задачи систем Data Mining.

Билет №10

Вопрос 1	Вопрос 2
Понятие «витрина данных»: назначение, достоинства и недостатки.	Стадии решения задач системами Data Mining: стадия свободного поиска, стадия анализа исключений.

Билет №11

Вопрос 1	Вопрос 2
Оптимизация витрин данных: таблицы итогов.	Классификация методов Data Mining по принципу работы с исходными обучающими данными.

Билет №12

Вопрос 1	Вопрос 2
Метод полного и инкрементального перевычисления для сопровождения таблиц итогов.	Классификация методов Data Mining, основанная на различных подходах к обучению математических моделей.

Билет №13

Вопрос 1	Вопрос 2
Процесс ETL: предназначение и основные классы.	Методы Data Mining: статистические, кибернетические.

Билет №14

Вопрос 1	Вопрос 2
Процесс ETL: стадия загрузки источника данных.	Основные задачи Data Mining: классификация.

Билет №15

Вопрос 1	Вопрос 2
Процесс ETL: стадия извлечения данных (IDC).	Основные задачи Data Mining: кластеризация.

Билет №16

Вопрос 1	Вопрос 2
Процесс ETL: стадия организации выгрузки данных.	Основные задачи Data Mining: прогнозирование.

Билет №17

Вопрос 1	Вопрос 2
Ошибки на стадии извлечения данных.	Основные задачи Data Mining: визуализация.

Билет №18

Вопрос 1	Вопрос 2
Стадия извлечения данных ETL: потоки STER и STAC.	Методы Data Mining: метод деревьев и его применение.

Билет №19

Вопрос 1	Вопрос 2
Процедура преобразования данных ETL: основные этапы.	Методы Data Mining: иерархические агломеративные и дивизимные методы, алгоритм к-средних для решения задач

	кластеризации.
--	----------------

Билет №20

Вопрос 1	Вопрос 2
Процесс ETL: три способа оптимизации очистки данных при загрузке.	Методы Data Mining: методы визуализации.

Билет №21

Вопрос 1	Вопрос 2
Проблемы, возникающие при работе с ХД, требующие оптимизации.	Data Mining: задачи анализа текстов.

Билет №22

Вопрос 1	Вопрос 2
Принципы ILM для Хранилищ данных.	Задачи анализа текстов: автоматическое построение тематических сюжетов.

Билет №23

Вопрос 1	Вопрос 2
Классификация проблем качества данных.	Автоматический анализ содержания документа с использованием семантической сети.

Билет №24

Вопрос 1	Вопрос 2
Проблема очистки данных на уровне схемы для единственного источника.	Задачи анализа текстов: технологии автоматического анализа содержания документа.

Билет №25

Вопрос 1	Вопрос 2
Основные системы хранения данных. Network Attached Storage (NAS).	Семантический разрыв и его основные причины.

Билет №26

Вопрос 1	Вопрос 2
Что такое системы оперативной обработки транзакций OLTP?	Отличие задач математической статистики от задач Data Mining.

Билет №27

Вопрос 1	Вопрос 2
Системы обработки гиперкубов: многомерная оперативная аналитическая обработка данных (MOLAP), реляционные системы (ROLAP), гибридная архитектура HOLAP.	Технология Data Mining: применение для решения основных задач.

Билет №28

Вопрос 1	Вопрос 2
Основные понятия многомерной базы данных: «факт», «измерение», «параметр измерения».	Семантические разрывы: вавилонский, кросспотоковый, кроссязыковый и асинхронный разрывы.

Лист регистрации изменений и дополнений

№ п/п	№ страницы внесения изменений	Дата внесения изменения	Краткое содержание изменений (основание)	Ф.И.О., подпись	«Согласовано» заве- дующий кафедрой, ведущей дисциплину