

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

 **УТВЕРЖДАЮ**
Декан факультета ВМК МГУ
/И.А.Соколов/
2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Хранилища данных. Анализ данных

Уровень высшего образования:

бакалавриат

Направление подготовки (специальность):

02.03.02 Фундаментальная информатика и информационные технологии

Направленность (профиль) ОПОП (ФИИТ):

дисциплина относится к вариативной части программы

Форма обучения:

очная

Москва 2023

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ бакалавриата, магистратуры, реализуемых последовательно по схеме интегрированной подготовки по направлениям 02.03.02, 02.04.02 «Фундаментальная информатика и информационные технологии» в редакции приказа МГУ от 30 декабря 2016 г.

Цель курса: получение теоретических знаний и практических навыков по проектированию и разработке хранилищ данных корпоративного уровня, разработка приложений для работы с такими хранилищами – задачи интеграции данных, система корпоративной отчетности, задачи интеллектуального анализа данных. Хранилища используют как реляционные транзакционные OLTP сервера данных, так и многомерные модели (кубы с иерархическими координатами) OLAP сервера; достижение уровня подготовки для сдачи сертификационного экзамена корпорации Microsoft для разработчиков OLAP и ETL решений, систем корпоративной отчетности. Курс изучается с использованием программных продуктов Microsoft SQL Server и Microsoft Visual Studio в среде SQL Server Data Tools (SSDT).

Курс состоит из пяти частей.

1. Первая часть знакомит студентов с понятиями иерархических хранилищ и аналитической обработкой данных на примере Microsoft SQL Server Analysis Services. Изучается универсальная модель данных для OLAP, иерархическое представление данных, основные требования к хранилищам по Кодду. Рассматривается логическая и физическая архитектура OLAP проекта, измерения и функции фактов (меры) в OLAP кубах, разработка кубов, базовой и расширенной функциональности для конечных пользователей. Рассматривается много технических деталей для разработки OLAP проектов – денормализация данных, схемы звезды, снежинки, родители-дети для координат, разрывы данных, иерархии данных по уровням, меры, зависимые и независимые координаты в кубе, кортежи, ячейки, гранулированность (атомарность) данных, агрегатные функции в кубе и хранение агрегированных данных в различном виде, динамика координат в жизненном цикле, статические и динамические элементы в кубе, процессинг и развертывание куба, вопросы оптимизации, производительности, безопасного доступа для многопользовательского режима.

2. Вторая часть курса посвящена изучению языка программирования MDX (Языка Многомерных Выражений – Multidimensional Expressions) для разработки OLAP кубов. Рассматриваются основы языка, выражения, функции и свойства, MDX вычисления на кубе и в измерениях, методы программирования расширенной функциональности в кубе.

3. Третья часть курса позволяет студентам научиться разрабатывать корпоративную отчетность для OLTP и OLAP проектов с помощью SQL Server Reporting Services. Рассматриваются требования к разработке отчетности, процесс разработки и развертывания отчетов на веб-сервер отчетности, управление отчетами, подписки, система безопасности.

4. Четвертая часть курса посвящена крайне важному классу задач интеграции разнородных информационных ресурсов данных. Изучается разработка проектов интеграции данных с помощью SQL Server Integration Services: технические, организационные, экономические требования к процессам интеграции данных, планирование проектов интеграции; потоки управления и потоки данных, контейнеры, пакеты, задачи потока данных, трансформации потока данных, ETL процессы, обработка ошибок в пакетах, ведение журналов.

5. В пятой части студентам дается понятие о системах интеллектуального анализа данных (Data Mining). Дается обзор типов решаемых задач (распознавания, классификации, прогноза) и обзор девяти алгоритмов Data Mining в Microsoft SQL Server. Один из алгоритмов – деревья решений – подробно рассматривается на примере маркетинговой задачи задач классификации и кластеризации групп пользователей для мультинациональной торгово - производственной компании с демонстрацией программного кода.

Практическая поддержка курса: семинарские занятия и лабораторные работы.

Семинарские и практические занятия

Каждую неделю проводится одно практическое занятие по курсу «Хранилища данных. Анализ данных».

Практическая часть курса поддержана практикумом – лабораторными работами по каждой теме. Все работы выполняются на компьютере в среде SQL Server Data Tools (SSDT). Теоретическая часть отрабатывается устно на семинарах.

Основные цели практикума:

- изучение теоретических понятий иерархических хранилищ и аналитической обработки данных;
- практическое изучение языка программирования MDX для разработки OLAP кубов;
- освоение теоретических и практически навыков в разработке корпоративной отчетности для OLTP и OLAP проектов с помощью SQL Server Reporting Services;
- практическое освоение задач интеграции разнородных информационных ресурсов данных;
- практическое освоение систем интеллектуального анализа данных;
- приобретение практического опыта работы с продуктами Microsoft SQL Server, Microsoft Visual Studio и средой SQL Server Data Tools.

Лабораторные работы по каждой теме студенты выполняют на компьютере в учебном классе и дома, показывают их на занятии преподавателю.

1. Дисциплина относится к вариативной части ОПОП ВО.

2. Входные требования для освоения дисциплины (модуля): учащиеся должны владеть знаниями по компьютерным и информационным наукам в объеме, соответствующем программе второго года обучения основных образовательных программ бакалавриата по укрупненным группам направлений и специальностей 02.00.00 «Компьютерные и информационные науки» (или отсутствуют). При изучении дисциплины используются знания, полученные при изучении дисциплины «Основы программирования» (2 семестр, блок общенаучной подготовки), а также знания, полученные в курсе «Разработка баз данных. Microsoft SQL Server» (5 семестр, блок профессиональной подготовки). Практические навыки, полученные при изучении дисциплины, могут быть использованы при выполнении квалификационной работы

3. Результаты обучения по дисциплине (модулю), соотношенные с требуемыми компетенциями выпускников.

3.3 Компетенции выпускников, освоивших программу **бакалавриата по направлению ФИИТ** частично формируемые при реализации дисциплины (модуля):

- ОПК-2. Способность применять в профессиональной деятельности современные языки программирования и языки баз данных, методологии системной инженерии, системы автоматизации проектирования, электронные библиотеки и коллекции, сетевые технологии, библиотеки и пакеты программ, современные профессиональные стандарты информационных технологий;
- ОПК-3. Способность 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;

3.4 Профессиональные компетенции выпускников, освоивших программу **бакалавриата по направлению ФИИТ** частично формируемые при реализации дисциплины (модуля):

научно-исследовательская деятельность:

- ПК-3. Способность использовать современные инструментальные и вычислительные средства;
- ПК-4. Способность решать задачи профессиональной деятельности в составе научно-исследовательского и производственного коллектива;
- ПК-6. Способность эффективно применять базовые математические знания и информационные технологии при решении проектно-технических и прикладных задач, связанных с развитием и использованием информационных технологий;

Планируемые результаты обучения по дисциплине (модулю):

Знать:

Раздел 1. Хранилища данных. Введение в OLAP.

1. История вопроса аналитической обработки данных. Решаемые задачи. Требования к OLAP системе для хранилища данных по Кодду. Тест FASMI. Планирование OLAP проекта: сбор и анализ бизнес требований, технических требований, выявление ограничений, уточнение модели и процессов. Демонстрация OLAP куба в Microsoft SQL Server.
2. Логическая архитектура OLAP проекта. Универсальная модель данных UDM (Unified Dimensional Model) в OLAP проекте. Планирование основных объектов OLAP куба – измерения и меры. Схемы связывания фактов и координат куба в хранилище – звезда, снежинка, иерархия.
3. Проектирование координат в кубе: атрибуты элементов (тип, ключ, имя, значение, сортировка, язык и др.), свойства атрибутов, индивидуализация наблюдателя, иерархия, сортировка элементов в координате по ключу, имени, значению элемента. Группировка и сортировка атрибутов. Меры в кубе.
4. Расширенные свойства атрибутов и измерений в кубе. Меняющиеся измерения – динамические, медленно меняющиеся и деформированные измерения. Операции в динамических измерениях. Ключ элемента и суррогатный ключ. Виртуальные измерения.
5. Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLONAP; кеширование – стандартные сценарии и пользовательские сценарии на XMLA (XML для аналитики). Секции и распределенные данные в реляционной СУБД. Секции в кубах. Вопросы производительности, администрирования, управления.
6. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарий XMLA, мастер развертывания, мастер синхронизации, автоматизация объектов АМО, создание резервной копии и ее восстановление. Система безопасности. Роли OLAP сервера и БД. Разрешения и их виды. Детализация разрешений для объектов куба – данных и процедур. Шифрование. Безопасность ОС и файловой системы. Производительность - разработка индексов, оптимизация запросов, оптимизация кеширования.

Раздел 2. Язык программирования MDX для OLAP.

7. Основы языка MDX. Оператор SELECT в MDX. Полные и неполные кортежи. Многомерные MDX-выражения для измерений, иерархий, элемента, множества, куба. MDX-операторы. MDX-функции измерений, иерархий, элементов, уровней, логические, числовые, строковые, кортежей, над множествами – навигация, сортировка, фильтры. Свойства элементов и ячеек.
8. MDX вычисления. Вычисления в кубе – вычисляемые члены в момент исполнения - операторы WITH...SELECT, CREATE MEMBER, MDX-сценарий с разным временем жизни; именованные наборы, вычисления в заданной области видимости – операторы SCOPE и THIS. Вычисления в измерениях – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.
9. Программирование на MDX расширенной функциональности в кубе. Ключевые индикаторы производительности – значение, цель, статус, индикаторы, тренд, вес. Действия – серверные MDX команды. Хранимые процедуры - .NET и COM сборки, регистрация их на сервере, назначение прав. Перспективы и переводы в кубе.

Раздел 3. Разработка корпоративной отчетности.

10. Система разработки отчетов SQL Server Reporting Services. Требования к системам отчетности. Планирование проекта отчетности. Демонстрация видов отчетов. Архитектура платформы отчетности - ASP.NET. Сервер отчетов, конструктор отчетов, конструктор модели, построитель динамических отчетов. Форматы отчетов. Функциональность отчетов. Диспетчер отчетов для управления отчетами – доступ, просмотр, подписки, защита, доставка отчетов. Планирование проекта отчетности.
11. Разработка и развертывание отчетов, управление отчетами и безопасность. Разработка отчетов. Разделение данных и представления, опубликованный отчет, процессированный отчет. Параметры в отчетах. Детализация drill down. Модель данных нерегламентированного отчета. Связанные отчеты. Моментальный снимок отчета. План разработки отчета. Табличные и матричные отчеты. Элементы отчетов – карта документа, таблицы, агрегированные данные, списки, выражения, диаграммы, рисунки, колоннитулы, номера страниц, форматирование. План тестирования отчета. Методы развертывания отчетов. Способы развертывания. Доступ и учетные записи. Консоль конфигурирования свойств. Управление отчетами. Виды поставляемых отчетов. Доставка отчетов. Виды доставки. Подписки. Система безопасности SQL Reporting Services.

Раздел 4. Проекты интеграции данных, ETL процессы.

12. Интеграция информационных ресурсов в хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services. Планирование проекта интеграции данных: экономические, организационные, технологические задачи. Источники и приемники данных. Оценка и проверка исходных данных. Извлечение данных. Адаптеры данных. Компоненты просмотра данных. Промежуточное хранение данных. Преобразование данных. Поддержка целостности. Загрузка в хранилище и киоски данных.

13. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи. Разработка пакетов - поток управления, поток данных, контейнеры данных, адаптеры данных, трансформации потока данных, виды трансформаций данных, обработка ошибок компонентов, ведение журналов, свойства пакетов, конфигурации пакетов. Примеры.

Раздел 5. Введение в задачи и алгоритмы Data Mining.

14. Введение в интеллектуальный анализ данных Data Mining в SQL Server. Типы решаемых задач – классификация, регрессия, кластеризация, ассоциативные правила, анализ последовательностей. Алгоритмы. Процесс разработки проекта Data Mining – подготовка данных, создание структуры Data Mining, выбор алгоритма, расчет модели, развертывание проекта, просмотр созданной модели с помощью визуализатора. Сценарии прогноза для анализа данных. Примеры.

15. Алгоритмы Data Mining. Интеграция OLAP, Integration Services, Reporting Services, Data Mining. Примеры.

Уметь:

1. Работать с Oracle VM VirtualBox. Ознакомиться с перечнем лабораторных работ в справочной документации SQL Server Analysis Services Tutorials. Установка баз данных для лабораторных работ – AdventureWorks и AdventureWorksDW, знакомство с их бизнес мо-

- дью. Освоение интерфейса работы с OLAP проектом в Microsoft Visual Studio в среде SQL Server Data Tools (SSDT). – компиляция, процессинг, развертывание и просмотр куба в браузере конструктора кубов для примера из Analysis Services Tutorial – AdventureWorksDW.
2. Логическая архитектура OLAP проекта. Планирование бизнес требований по интервью с четырьмя сотрудниками заказчика из компании. AdventureWorksDW (этап 1, файлы выдает преподаватель). Проектирование координат куба со структурой звезды, снежинка, иерархия, Лаб. работа из Module 2 (PPT) - разработка первого OLAP-куба.
 3. Проектирование измерений и атрибутов. Атрибуты элементов и их свойства (тип, ключ, имя, значение, сортировка, язык и др.), переименование атрибутов, удаление лишних атрибутов, создание новых атрибутов - именованных вычислений. Группировка атрибутов в папки. Составные ключи атрибутов для измерения Time. Сортировка по ключу, значению элемента. Создание иерархий. Меры в кубе. Форматирование мер. Лаб. работа из Analysis Services Tutorial №2,3 – «Определение и развертывание куба» - базовая функциональность.
 4. Определение расширенных свойств атрибутов и измерений в кубе. Лаб. работа №4 из Tutorial. Связанные атрибуты, фиксированные и гибкие связи. Расширенная сортировка по ключу, значению атрибута или по вторичному (связанному) атрибуту. Автоматическое группирование элементов атрибута. Видимые и невидимые элементы в измерении родитель-потомок. Значения по умолчанию и обработка значений NULL. Лаб. работа Tutorial №4.
 5. Меняющиеся измерения – динамические, медленно меняющиеся и деформированные измерения. Операции, свертки и функции в динамических элементах. Связи между атрибутами фиксированные и гибкие. Ключ элемента и суррогатный ключ. Виртуальные измерения. Лаб. работа из Module 3 (PPT).
 6. Определение связей между измерениями. Определение ссылочной связи с помощью связи «первичный ключ — внешний ключ». Измерения на основе таблицы фактов. Определение связи измерений как связи фактов. Определение связи «многие ко многим» между таблицами измерений и таблицами фактов. Задание степени granularity измерения в группе мер. Лаб. работа №5 из Tutorial.
 7. Планирование физической архитектуры OLAP проекта по спецификациям заказчика - лаб. работа из Module 4 (PPT).
 8. Развертывание OLAP проектов интерактивно и при помощи скрипта. Основные задачи администрирования OLAP проекта. Планирование прав и назначение прав пользователей на объекты безопасности в кубе. Проверка заданных разрешений. Лаб. работа из Module 9 (PPT) – Развертывание проекта, лаб. работа №10 из Tutorial – назначение прав на объекты в кубе.
 9. MDX-выражения, функции и свойства. Оператор SELECT. Вычисления в измерениях и в мерах. Вычисляемые элементы, именованные наборы и скрипты с различной областью видимости. Операторы SCOPE и THIS. Лаб. работы из Module 6 (PPT) и Tutorial №6.
 10. Программирование расширенной функциональности в кубе. Ключевые индикаторы производительности – значение, цель, статус, индикаторы, тренд, вес. Действия – серверные MDX команды. Перспективы и переводы. Лаб. работа из Tutorial №7,8,9.

11. Система разработки отчетов SQL Server Reporting Services. Планирование проекта отчетности – Лаб. работа из Module 10. Разработка отчетов с базовой функциональностью – табличный отчет. Лаб. работа из Tutorial «Основной отчет. Группировка, сортировка, форматирование и стиль, подитого».
12. Разработка отчетов с расширенной функциональностью. Табличные отчеты. Лаб. работа из Tutorial «Динамические запросы в отчетах, параметры, списки, компоненты (календарь)».
13. Разработка и развертывание отчетов, управление отчетами и безопасностью. Лаб. работа из Tutorial “Создание управляемой данными подписки”. Лаб. работа из Module 10 (PPT) “Настройка конфигураций, ролей безопасности и разрешений на сервере отчетности”.
14. Задачи интеграции данных. SQL Server Integration Services. Разработка пакетов. Лаб. работы из Tutorial «Разработка простого ETL пакета», «Добавление циклов», знакомство с кодом пакета «Обработка XML данных».
15. Разработка пакетов. Лаб. работы из Tutorial «Конфигурация пакетов», «Журналы», «Перенаправление потока ошибок».
16. Интеграция OLAP и ETL задач. Знакомство с образцами пакетов «Data Cleaning» - очистка данных с алгоритмом нечеткой логики, «Calculated Columns» - пакет преобразования данных, «SCD_DataLoading» - разработка медленно меняющегося измерения второго типа Customers в OLAP кубе, загрузка в него данных, пакет обновления данных в SCD измерения куба при смене фамилии клиента, визуализация в кубе разных фамилий клиент на разные даты; «AWData WarehouseRefresh» - ETL процессы - извлечение данных из реляционной БД, использование промежуточного хранения, создание и заполнение таблиц в новой БД, установка в ней ключей и ограничений, заполнение данными.

Владеть:

1. Навыками разработки физической архитектуры OLAP проекта.
2. Навыками программирования на языке MDX расширенной функциональности в кубе.
3. Навыками разработки отчетов SQL Server Reporting Services.
4. Формат обучения: для проведения лекций-презентаций используется медиапроектор и компьютер, для проведения практикума используется компьютерный класс. Для самостоятельной работы дома среда разработки Microsoft SQL Server и Microsoft Visual Studio предоставляет студентам в рамках международной программы DreamSpark Premium.
5. Объем дисциплины (модуля) составляет 3 зачетные единицы (108 часов), в том числе 72 академических часа, отведенных на контактную работу обучающихся с преподавателем (лекции – 36 часов, практикум – 36 часов), самостоятельная работа - 36 часов. Зачёт в 6 семестре.
6. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий.

Наименование и краткое содержание разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе			
		Контактная работа (работа во взаимодействии с преподавателем)		Самостоятельная работа обучающегося, часы	
		Виды контактной работы, часы	Занятия семинарского типа	Всего	
		*Занятия лекционные	*Занятия семинарского типа		
1. Хранилища данных. OLAP проекты	32	12	12	24	8
2. Язык программирования MDX для OLAP	18	6	6	12	6
3. Разработка корпоративной отчетности	20	6	6	12	8
4. Проекты интеграции данных, ETL процессы	18	6	6	12	6
5. Введение в задачи и алгоритмы Data Mining	16	6	6	12	4
Промежуточная аттестация: зачёт	4	0	0	0	4
Итого.	108	36	36	72	36

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)

7.1. Контроль преподавателем выполнения лабораторных работ с выставлением рейтинговых баллов для проведения текущего контроля успеваемости

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

Вопросы к зачёту.

1. Что такое OLAP. Характеристики OLAP по Э.Кодду. Тест FASMI.
2. Понятие хранилища данных.
3. Понятие и модель данных OLAP.
4. Структура OLAP-куба.

5. Иерархия измерений OLAP-кубов.
6. Операции, выполняемые над гиперкубом.
7. Таблица фактов.
8. Таблицы измерений.
9. Архитектура OLAP-систем.
10. Клиентские OLAP-средства.
11. Серверные OLAP-средства.
12. Технические аспекты многомерного хранения данных.
13. Возможности службы SSAS.
14. Инструменты управления службой SSAS.
15. Логическая архитектура SSAS.
16. Физическая архитектура . SSAS.
17. Архитектура программирования SSAS.
18. Проектирование и реализация многомерных баз данных с использованием SSAS.
19. Запросы к многомерным базам данных с использованием языка программирования MDX.
20. Синтаксис базовой инструкции SELECT языка программирования MDX с использованием предложений SELECT, FROM и WHERE.
21. Использование наборов в запросах языка программирования MDX.
22. Возможности Integration Services для работы с OLAP.
23. Архитектура служб SSIS.
24. Пакет SSIS.
25. Планирование развертывания служб Analysis Services.
26. Настройка безопасности служб Analysis Services.
27. Система разработки отчетов Reporting Services. Требования к решению создания отчетов.
28. Создание отчетов Reporting Services.
29. Размещение отчетов Reporting Services.
30. Доступ к отчетам Reporting Services.
31. Доставка отчетов Reporting Services.
32. Дополнительные возможности служб Reporting Services.
33. Администрирование служб Reporting Services.

34. Упрощенные механизмы системы отчетности Reporting Services.
35. Что следует учитывать при проектировании отчета Reporting Services.
36. Архитектура служб Reporting Services: компоненты сервера.
37. Архитектура служб Reporting Services: клиентские компоненты.
38. Масштабируемость Reporting Services.
39. Классификация задач Data Mining.
40. Практическое применение Data Mining.
41. Алгоритмы Data Mining: 1-Rule алгоритм, алгоритм Naive Bayes. Представление результатов в виде дерева решений.

Билет для итогового зачёта состоит из двух вопросов, например

1. Операции, выполняемые над гиперкубом
2. Классификация задач Data Mining.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации самостоятельной работы студентов и рубежного контроля.

В течение семестра студенты выполняют 16 лабораторных работ по 5 баллов за работу. В конце семестра студенты сдают зачёт с двумя теоретическими вопросами.

Эффективность освоения курса оценивается по рейтинговой системе. Максимальный балл – 100.

Таблица. Вклад в рейтинговую оценку разных учебных заданий

№	Учебные задания	Баллы
1	16 лабораторных работ, по 5 баллов за работу	80 баллов
2	Итоговый зачёт. Два теоретических вопроса	20 баллов

Оценка по курсу выставляется, исходя из следующих критериев:

Оценка	Минимальное количество баллов	Максимальное количество баллов
ЗАЧЕТ	60	100
НЕЗАЧЕТ	0	59

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)				
Оценка	2	3	4	5
РО и соответствующие виды оценочных средств				
Знания <i>Зачёт</i>	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения <i>Контроль выполнения лабораторных работ с выставлением рейтинговых баллов</i>	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности принципиального характера)	Успешное и систематическое умение
Навыки <i>(владения, опыт деятельности)</i> <i>Контроль выполнения лабораторных работ с выставлением рейтинговых баллов</i>	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

Соответствие результатов обучения и компетенций, в развитии которых участвует дисциплина (модуль)	
Результаты обучения	Компетенция, с которой формируем достижимые результаты обучения
<p>Знать:</p> <p>1. Основы языка MDX. Оператор SELECT в MDX. Полные и неполные кортежи. Многомерные MDX-выражения для измерений, иерархии, элемента, множества, куба. MDX-операторы. MDX-функции измерений, иерархий, элементов, уровней, логические, числовые, строковые, кортежей, над множествами – навигация, сортировка, фильтры. Свойства элементов и ячеек.</p> <p>2. MDX вычисления. Вычисления в кубе – вычисляемые члены в момент исполнения - операторы WITH...SELECT, CREATE MEMBER, MDX-сценарий с разным временем жизни; именованные наборы, вычисления в заданной области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вы-</p>	ОПК-2.

	<p>числяемые ячейки. Изменение порядка прохода ячеек.</p> <p>3. Программирование на MDX расширенной функциональности в кубе. Ключевые индикаторы производительности – значение, цель, статус, индикаторы, тренд, вес. Действия – серверные MDX команды. Хранимые процедуры - .NET и COM сборки, регистрация их на сервере, назначение прав. Перспективы и переводы в кубе.</p> <p>Уметь:</p> <ol style="list-style-type: none"> 1. MDX-выражения, функции и свойства. Оператор SELECT. Вычисления в измерениях и в мерах. Вычисляемые элементы, именованные наборы и скрипты с различной областью видимости. Операторы SCOPE и THIS. Лаб.работы из Module 6 (PPT) и Tutorial №6. 2. Программирование расширенной функциональности в кубе. Ключевые индикаторы производительности – значение, цель, статус, индикаторы, тренд, вес. Действия – серверные MDX команды. Перспективы и переводы. Лаб.работа из Tutorial №7,8,9. <p>Владеть:</p> <ol style="list-style-type: none"> 1. навыками программирования на языке MDX расширенной функциональности в кубе
<p>ОПК-3.</p>	<p>Знать:</p> <ol style="list-style-type: none"> 1. История вопроса аналитической обработки данных. Решаемые задачи. Требования к OLAP системе для хранилища данных по Кодду. Тест FASMI. Планирование OLAP проекта: сбор и анализ бизнес требований, технических требований, выявление ограничений, уточнение модели и процессов. Демонстрация OLAP куба в Microsoft SQL Server. 2. Логическая архитектура OLAP проекта. Универсальная модель данных UDM (Unified Dimensional Model) в OLAP проекте. Планирование основных объектов OLAP куба – измерения и меры. Схемы связывания фактов и координат куба в хранилище – звезда, снежинка, иерархия. 3. Проектирование координат в кубе: атрибуты элементов (тип, ключ, имя, значение, сортировка, язык и др.), свойства атрибутов, индивидуализация наблюдателя, иерархия, сортировка элементов в координате по ключу, имени, значению элемента. Группировка и сортировка атрибутов. Меры в кубе. 4. Расширенные свойства атрибутов и измерений в кубе. Меняющиеся измерения – динамические, медленно меняющиеся и деформированные измерения. Операции в динамических измерениях. Ключ элемента и суррогатный ключ. Виртуальные измерения. 5. Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HОLAP; кеширование – стандартные сценарии и пользовательские сценарии на XMLA (XML для аналитики). Секции и распределенные данные в реляционной СУБД. Секции в кубах. Вопросы производительности, администрирования, управления. <p>Уметь:</p>

	<p>1. Работать с Oracle VM VirtualBox. Ознакомиться с перечнем лабораторных работ в справочной документации SQL Server Analysis Services Tutorials. Установка баз данных для лабораторных работ – AdventureWorks и AdventureWorksDW, знакомство с их бизнес моделью. Освоение интерфейса работы с OLAP проектом в Microsoft Visual Studio в среде SQL Server Data Tools (SSDT). – компиляция, процессинг, развертывание и просмотр куба в браузере конструктора кубов для примера из Analysis Services Tutorial – AdventureWorksDW.</p> <p>2. Логическая архитектура OLAP проекта. Планирование бизнес требований по интервью с четырьмя сотрудниками заказчика из компании. AdventureWorksDW (этап 1, файлы выдает преподаватель). Проектирование координат куба со структурой звезда, снежинка, иерархия, Лаб. работа из Module 2 (PPT) - разработка первого OLAP-куба.</p> <p>3. Проектирование измерений и атрибутов. Атрибуты элементов и их свойства (тип, ключ, имя, значение, сортировка, язык и др.), переименование атрибутов, удаление лишних атрибутов, создание новых атрибутов - именованных вычислений. Группировка атрибутов в папки. Составные ключи атрибутов для измерения Time. Сортировка по ключу, значению элемента. Создание иерархий. Меры в кубе. Форматирование мер. Лаб. работа из Analysis Services Tutorial №2,3 – «Определение и развертывание куба» - базовая функциональность.</p> <p>4. Определение расширенных свойств атрибутов и измерений в кубе. Лаб. работа №4 из Tutorial. Связанные атрибуты, фиксированные и гибкие связи. Расширенная сортировка по ключу, значению атрибута или по вторичному (связанному) атрибуту. Автоматическое группирование элементов атрибута. Видимые и невидимые элементы в измерении родитель-потомок. Значения по умолчанию и обработка значений NULL. Лаб. работа Tutorial №4.</p> <p>5. Меняющиеся измерения – динамические, медленно меняющиеся и деформированные измерения. Операции, свертки и функции в динамических элементах. Связи между атрибутами фиксированные и гибкие. Ключ элемента и суррогатный ключ. Виртуальные измерения. Лаб. работа из Module 3 (PPT).</p> <p>6. Определение связей между измерениями. Определение ссылочной связи с помощью связи «первичный ключ — внешний ключ». Измерения на основе таблицы фактов. Определение связи измерений как связи фактов. Определение связи «многие ко многим» между таблицами измерений и таблицами фактов. Задание степени granularity измерения в группе мер. Лаб. работа №5 из Tutorial.</p> <p>7. Планирование физической архитектуры OLAP проекта по спецификациям заказчика - лаб. работа из Module 4 (PPT).</p> <p>Владеть:</p>	
	<p>1. навыками разработки физической архитектуры OLAP проекта</p> <p>Знать:</p>	<p>1. Система разработки отчетов SQL Server Reporting Services. Требования к системам отчетности. Планирование проекта отчетности. Демонстрация видов отчетов. Архитектура платформы отчетности - ASP.NET. Сервер отчетов, конструктор отчетов, конструктор модели, построитель динамических отчетов. Форматы отчетов. Функциональность отчетов. Диспетчер отчетов для управления отчетами – доступ, просмотр, подписки, защита, доставка отчетов.</p> <p>ПК-3.</p>

	<p>тов. Планирование проекта отчетности.</p> <p>2. Разработка и развертывание отчетов, управление отчетами и безопасностью. Разработка отчетов. Разделение данных и представления, опубликованный отчет, процессированный отчет. Параметры в отчетах. Детализация drill down. Модель данных нерегламентированного отчета. Связанные отчеты. Моментальный снимок отчета. План разработки отчета. Табличные и матричные отчеты. Элементы отчетов – карта документа, таблицы, агрегированные данные, списки, выражения, диаграммы, рисунки, колоннитулы, номера страниц, форматирование. План тестирования отчета. Методы развертывания отчетов. Способы развертывания. Доступ и учетные записи. Консоль конфигурирования свойств. Управление отчетами. Виды поставляемых отчетов. Доставка отчетов. Виды доставки. Подписки. Система безопасности SQL Reporting Services.</p> <p>Уметь:</p> <ol style="list-style-type: none"> 1. Система разработки отчетов SQL Server Reporting Services. Планирование проекта отчетности – Лаб. работа из Module 10. Разработка отчетов с базовой функциональностью – табличный отчет. Лаб. работа из Tutorial «Основной отчет. Группировка, сортировка, форматирование и стиль, подитоги». 2. Разработка отчетов с расширенной функциональностью. Табличные отчеты. Лаб. работа из Tutorial «Динамические запросы в отчетах, параметры, списки, компоненты (календарь)». 3. Разработка и развертывание отчетов, управление отчетами и безопасностью. Лаб. работа из Tutorial “Создание управляемой данными подписки”. Лаб. работа из Module 10 (PPT) “Настройка конфигураций, ролей безопасности и решений на сервере отчетности”. <p>Владеть:</p> <ol style="list-style-type: none"> 1. навыками разработки отчетов SQL Server Reporting Services 	
	<p>Знать:</p> <ol style="list-style-type: none"> 1. Интеграция информационных ресурсов в хранилища данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services. Планирование проекта интеграции данных: экономические, организационные, технологические задачи. Источники и приемники данных. Оценка и проверка исходных данных. Извлечение данных. Адаптеры данных. Компоненты просмотра данных. Промежуточное хранение данных. Преобразование данных. Поддержка целостности. Загрузка в хранилище и киоски данных. 2. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи. Разработка пакетов - поток управления, поток данных, контейнеры данных, адаптеры данных, трансформации потока данных, виды трансформаций данных, обработка ошибок компонентов, ведение журналов, свойства пакетов, конфигурации пакетов. Примеры. 3. Введение в интеллектуальный анализ данных Data Mining в SQL Server. Типы решаемых задач – классификация, регрессия, кластеризация, ассоциативные правила, анализ последовательностей. Алгоритмы. Процесс разработки проекта Data Mining – подготовка данных, создание структуры Data Mining, выбор алгоритма, расчет модели, раз- 	<p>ПК-4</p>

<p>вертывание проекта, просмотр созданной модели с помощью визуализатора. Сценарии прогноза для анализа данных. Примеры.</p> <p>4. Алгоритмы Data Mining. Интеграция OLAP, Integration Services, Reporting Services, Data Mining. Примеры.</p> <p>Уметь:</p> <ol style="list-style-type: none"> 1. Задачи интеграции данных. SQL Server Integration Services. Разработка пакетов. Лаб. работы из Tutorial «Разработка простого ETL пакета», «Добавление циклов», знакомство с кодом пакета «Обработка XML данных». 2. Разработка пакетов. Лаб. работы из Tutorial «Конфигурация пакетов», «Журналы», «Перенаправление потока ошибок». 3. Интеграция OLAP и ETL задач. Знакомство с образцами пакетов «Data Cleaning» - очистка данных с алгоритмом нечеткой логики, «Calculated Columns» - пакет преобразования данных, «SCD_DataLoading» - разработка медленно меняющегося измерения второго типа Customers в OLAP кубе, загрузка в него данных, пакет обновления данных в SCD измерении куба при смене фамилии клиента, визуализация в кубе разных фамилий клиента на разные даты; «AWData Warehouse Refresh» - ETL процессы - извлечение данных из реляционной БД, использование промежуточного хранения, создание и заполнение таблиц в новой БД, установка в ней ключей и ограничений, заполнение данными. 	<p>ПК-6.</p>
<p>Знать:</p> <ol style="list-style-type: none"> 1. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарий XMLA, мастер развертывания, мастер синхронизации, автоматизация объектов АМО, создание резервной копии и ее восстановление. Система безопасности. Роли OLAP сервера и БД. Разрешения и их виды. Детализация разрешений для объектов куба – данных и процедур. Шифрование. Безопасность ОС и файловой системы. Производительность - разработка индексов, оптимизация запросов, оптимизация кеширования. <p>Уметь:</p> <ol style="list-style-type: none"> 1. Развертывание OLAP проектов интерактивно и при помощи скрипта. Основные задачи администрирования OLAP проекта. Планирование прав и назначение прав пользователей на объекты безопасности в кубе. Проверка заданных разрешений. Лаб. работа из Module 9 (PPT) – Развертывание проекта, лаб. работа №10 из Tutorial – назначение прав на объекты в кубе. 	

8. Ресурсное обеспечение:

Основная литература

1. Microsoft SQL Server 2012. Создание запросов. Учебный курс Microsoft: Пер. с англ. / И. Бен-Ган, Д. Сарка, Р. Талмейдж. — М.: Издательство «Русская редакция», 2014.

2. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С.Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2007.
3. Разработка приложений на основе Microsoft SQL Server 2008 / Лобел Л., Браст Э.Дж., Форте С. : Пер. с англ. М.: Русская редакция; СПб.: БХВ-Петербург. 2010.
4. Microsoft SQL Server 2008. Разработка баз данных. Учебный курс Microsoft/ Тернстрем Т., Вебер Э., Хотек М. М.: Русская редакция; СПб.: БХВ-Петербург. 2010.
5. Виейра Р. Программирование баз данных Microsoft SQL Server 2005 для профессионалов. М.: Вильямс. 2008.
6. Нильсен Пол. SQL Server 2005. Библия пользователя. М.: Вильямс. 2008.
7. Codd E.F., Codd S.B., Salley, C.T.: Providing OLAP (On-Line Analytical Processing) to user-analysts: An IT mandate. Technical report. 1993. (Требования к разработке OLAP систем)
8. Codd E.F. A Relational Model of Data For Large Shared Data Banks. Communications of the Association for Computing Machinery. Vol. 13. No. 6. June 1970. pp. 377–387. (Реляционная модель данных для крупных, совместно используемых банков данных).
9. SQL Server Books OnLine – электронная документация к продукту.

Дополнительная литература

1. Барсегян А.А., Куприянов М.С., Степаненко В.В. и др. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург. 2004.
2. Райордан Р. Основы реляционных баз данных. М.: Microsoft Press. 2001.
3. Материалы по OLAP www.olarp.ru
4. Материалы по OLAP www.olarpreport.com
5. MSDN: Microsoft SQL Server. Technical Articles. Paul Sanders. Introduction to the Unified Dimensional Model (UDM) msdn.microsoft.com
6. MSDN: Microsoft SQL Server. Technical Articles. Richard Tkachuk. Many-to- Many Dimensions in Analysis Services 2005. msdn.microsoft.com
7. MSDN: Microsoft SQL Server. Analysis Services Technical Articles. T. K. Anand. Processing Architecture msdn.microsoft.com
8. Материалы сайтов по Data Mining www.kdnuggets.com, www.iso.ru, www.basegroup.ru, www.hyperion.ru, www.cognos.com, www.statsoft.com, <http://msdn.microsoft.com>, www.sqlserverdatamining.com/DMcommunity (Microsoft)

Материально-техническое обеспечение:

Медиапроектор и компьютер для проведения лекций-презентаций, компьютерный класс для проведения практикума. Для самостоятельной работы дома среда разработки Microsoft SQL Server и Microsoft Visual Studio предоставляется студентам в рамках международной программы DreamSpark Premium.

9. Язык преподавания - русский.

10. Преподаватель:

ст. н.с. факультета ВМК МГУ, к.ф.-м.н., *Microsoft CERTIFIED Professional* И.С.Барашков, *MCP ID# 3132306*.

11. Автор программы:

ст. н.с. факультета ВМК МГУ, к.ф.-м.н., *Microsoft CERTIFIED Professional* И.С.Барашков, *MCP ID# 3132306*.