

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Сыров Игорь Анатольевич  
Должность: Директор  
Дата подписания: 04.09.2023 11:28:53  
Уникальный программный ключ:  
b683afe664d7e9f64175886cf9626a196149ad36

СТЕРЛИТАМАКСКИЙ ФИЛИАЛ  
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО  
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ  
«УФИМСКИЙ УНИВЕРСИТЕТ НАУКИ И ТЕХНОЛОГИЙ»

Факультет Математики и информационных технологий  
Кафедра Математического моделирования

**Рабочая программа дисциплины (модуля)**

дисциплина ***Б1.О.25 Распределенные базы и хранилища данных***

обязательная часть

Направление

**01.03.02** ***Прикладная математика и информатика***  
код наименование направления

Программа

***Искусственный интеллект и анализ данных***

Форма обучения

**Очная**

Для поступивших на обучение в  
**2023 г.**

Разработчик (составитель)  
***кандидат химических наук, доцент кафедры математического моделирования***  
***Иремадзе Э. О.***  
ученая степень, должность, ФИО

<b>1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций .....</b>	<b>3</b>
<b>2. Цели и место дисциплины (модуля) в структуре образовательной программы .....</b>	<b>3</b>
<b>3. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических или астрономических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся .....</b>	<b>4</b>
<b>4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий.....</b>	<b>4</b>
4.1. Разделы дисциплины и трудоемкость по видам учебных занятий (в академических часах).....	4
4.2. Содержание дисциплины, структурированное по разделам (темам) .....	7
<b>5. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине (модулю).....</b>	<b>14</b>
<b>6. Учебно-методическое и информационное обеспечение дисциплины (модуля) .....</b>	<b>22</b>
6.1. Перечень учебной литературы, необходимой для освоения дисциплины (модуля) .....	22
6.2. Перечень электронных библиотечных систем, современных профессиональных баз данных и информационных справочных систем .....	23
6.3. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства .....	24
<b>7. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю) .....</b>	<b>24</b>

**1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций**

<b>Формируемая компетенция (с указанием кода)</b>	<b>Код и наименование индикатора достижения компетенции</b>	<b>Результаты обучения по дисциплине (модулю)</b>
ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-4.1. знать и понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства при решении задач профессиональной деятельности	Обучающийся должен: основные математические методы современной теории управления.
	ОПК-4.2. уметь выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности	Обучающийся должен: формализовать постановки прикладных задач теории автоматического управления, оценивать эффективность полученных решений.
	ОПК-4.3. иметь практический опыт применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Обучающийся должен: основными понятиями математического аппарата теории автоматического управления.

**2. Цели и место дисциплины (модуля) в структуре образовательной программы**

Цели изучения дисциплины:

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).

Дисциплина изучается на 2 курсе в 4 семестре

**3. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических или астрономических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся**

Общая трудоемкость (объем) дисциплины составляет 4 зач. ед., 144 акад. ч.

Объем дисциплины	Всего часов
	Очная форма обучения
Общая трудоемкость дисциплины	144
Учебных часов на контактную работу с преподавателем:	
лекций	16
практических (семинарских)	24
лабораторных	24
другие формы контактной работы (ФКР)	0,2
Учебных часов на контроль (включая часы подготовки):	
дифференцированный зачет	
Учебных часов на самостоятельную работу обучающихся (СР)	79,8

Формы контроля	Семестры
дифференцированный зачет	4

**4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий**

**4.1. Разделы дисциплины и трудоемкость по видам учебных занятий (в академических часах)**

№ п/п	Наименование раздела / темы дисциплины	Виды учебных занятий, включая самостоятельную работу обучающихся и трудоемкость (в часах)			СР
		Контактная работа с преподавателем			
		Лек	Пр/Сем	Лаб	

<b>1</b>	<b>Хранилища данных. OLAP проекты</b>	<b>8</b>	<b>10</b>	<b>4</b>	<b>30</b>
1.1	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.	0	2	0	8
1.2	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.	2	2	2	8
1.3	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.	2	2	2	4
1.4	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.	2	2	0	4
1.5	Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLAP; 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. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарии 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.	2	2	0	6
<b>2</b>	<b>Язык программирования MDX для OLAP</b>	<b>4</b>	<b>4</b>	<b>6</b>	<b>19,8</b>
2.1	Основы языка MDX. Оператор SELECT в MDX.	2	2	2	6

	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.				
2.2	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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.	2	2	2	6
2.3	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.	0	0	2	7,8
<b>3</b>	<b>Разработка корпоративной отчетности</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>12</b>
3.1	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. Планирование проекта отчетности.	2	0	2	4
3.2	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, колонтитулы, номера страниц, форматирование.	0	2	2	4
3.3	План тестирования отчета. Методы развертывания отчетов. Способы развертывания.	0	2	2	4

	Доступ и учетные записи. Консоль конфигурирования свойств. Управление отчетами. Виды поставляемых отчетов. Доставка отчетов. Виды доставки. Подписки. Система безопасности SQL Reporting Services.				
<b>4</b>	<b>Проекты интеграции данных, ETL процессы</b>	<b>2</b>	<b>6</b>	<b>8</b>	<b>18</b>
4.1	Интеграция информационных ресурсов в хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services.	0	2	2	4
4.2	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. Преобразования данных. Поддержка целостности. Загрузка в хранилище и киоски данных. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи.	0	2	2	4
4.3	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	2	2	4	10
	<b>Итого</b>	<b>16</b>	<b>24</b>	<b>24</b>	<b>79,8</b>

#### 4.2. Содержание дисциплины, структурированное по разделам (темам)

Курс практических/семинарских занятий

№	Наименование раздела / темы дисциплины	Содержание
<b>1</b>	<b>Хранилища данных. OLAP проекты</b>	
1.1	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.	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.
1.2	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	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.	куба в хранилище – звезда, снежинка, иерархия.
1.3	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.	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.
1.4	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.	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.
1.5	Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLAP; 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. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарии 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.	Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLAP; 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. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарии 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.
<b>2</b>	<b>Язык программирования MDX для OLAP</b>	
2.1	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,	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.	числовые, строковые, кортежей, над множествами – навигация, сортировка, фильтры. Свойства элементов и ячеек.
2.2	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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.	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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.
<b>3</b>	<b>Разработка корпоративной отчетности</b>	
3.2	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, колонтитулы, номера страниц, форматирование.	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, колонтитулы, номера страниц, форматирование.
3.3	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.	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.
<b>4</b>	<b>Проекты интеграции данных, ETL процессы</b>	
4.1	Интеграция информационных ресурсов в хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services.	Интеграция информационных ресурсов в хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services.
4.2	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.	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.

	Преобразования данных. Поддержка целостности. Загрузка в хранилище и киоски данных. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи.	Преобразования данных. Поддержка целостности. Загрузка в хранилище и киоски данных. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи.
4.3	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	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

#### Курс лабораторных занятий

№	Наименование раздела / темы дисциплины	Содержание
<b>1</b>	<b>Хранилища данных. OLAP проекты</b>	
1.2	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.	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.
1.3	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.	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.
<b>2</b>	<b>Язык программирования MDX для OLAP</b>	
2.1	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.	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.
2.2	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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении –	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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении –

	унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.	унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.
2.3	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.	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.
<b>3</b>	<b>Разработка корпоративной отчетности</b>	
3.1	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. Планирование проекта отчетности.	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.
3.2	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, колонтитулы, номера страниц, форматирование.	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, колонтитулы, номера страниц, форматирование.
3.3	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.	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.
<b>4</b>	<b>Проекты интеграции данных, ETL процессы</b>	
4.1	Интеграция информационных ресурсов в	Интеграция информационных ресурсов в

	хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services.	хранилищах данных. Типы решаемых задач. Жизненный цикл данных. ETL процессы. Архитектура SQL Server Integration Services.
4.2	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. Преобразования данных. Поддержка целостности. Загрузка в хранилище и киоски данных. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи.	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. Преобразования данных. Поддержка целостности. Загрузка в хранилище и киоски данных. Разработка пакетов для интеграции данных. Пакеты. Планирование разработки пакетов, решаемые задачи.
4.3	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	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

#### Курс лекционных занятий

№	Наименование раздела / темы дисциплины	Содержание
<b>1</b>	<b>Хранилища данных. OLAP проекты</b>	
1.2	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.	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.
1.3	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.	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.
1.4	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	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.	измерения.
1.5	<p>Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLAP; 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. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарии 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>	<p>Физическая архитектура OLAP проекта. Физический уровень куба – секции, выбор типа хранения данных - MOLAP, ROLAP, HOLAP; 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. Развертывание OLAP проектов. Система безопасности. Оптимизация производительности. Методы развертывания – интерактивный, сценарии 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>
<b>2</b>	<b>Язык программирования MDX для OLAP</b>	
2.1	<p>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>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>
2.2	<p>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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.</p>	<p>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 в области видимости – операторы SCOPE и THIS. Вычисления в измерении – унарные операторы, функции-члены. Вычисляемые ячейки. Изменение порядка прохода ячеек.</p>
<b>3</b>	<b>Разработка корпоративной отчетности</b>	
3.1	Система разработки отчетов SQL Server	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.	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.
<b>4</b>	<b>Проекты интеграции данных, ETL процессы</b>	
4.3	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	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

### 5. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа студентов, предусмотренная учебным планом, должна соответствовать более глубокому усвоению изучаемого материала, формировать навыки исследовательской работы и ориентировать их на умение применять полученные теоретические знания на практике. В процессе этой деятельности решаются задачи:

- научить студентов работать с учебной литературой;
- формировать у них соответствующие знания, умения и навыки;
- стимулировать профессиональный рост студентов, воспитывать творческую активность и инициативу.

Самостоятельная работа студентов предполагает:

- подготовку к занятиям (изучение лекционного материала и чтение литературы);
- оформление отчета по самостоятельной работе;
- подготовку к итоговому контролю.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам лекций, учебной и научной литературе) подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на лекциях и лабораторных занятиях,
- подготовку к лабораторным занятиям.

Обязательным является выполнение лабораторных работ, которые оформляются в специально отведённой для этого тетради и систематически сдаются на проверку.

Текущий контроль осуществляется в формах:

- опрос студентов;
- домашние работы;
- самостоятельная работа студентов на лабораторных занятиях.

Лабораторная работа № 1. Создание хранилища данных

Цель работы – подготовка реляционного хранилища данных, отражающего факты функционирования доменного цеха металлургического предприятия, средствами службы SQL Server Integration Services (SSIS).

#### Основные теоретические сведения

В условиях высокой конкуренции в сфере бизнеса, отсутствия стабильности экономических условий, в которых функционируют современные предприятия, устанавливаются все более жесткие требования к качеству принимаемых руководством предприятия решений на всех уровнях иерархии управления. Процесс принятия решений предполагает знания и учета множества аспектов и данных, характеризующих актуальное состояние и тенденции развития компании. В то же время объем этой информации ежедневно увеличивается. Эти факторы способствуют активному развитию систем бизнес-аналитики (Business Intelligence). В русскоязычных источниках также можно встретить аналогичный термин «системы поддержки принятия решений».

Под термином «Business Intelligence (BI)», или «бизнес-аналитика», понимают:

непосредственно процесс извлечения из накопленных в организации данных ранее неизвестной многоаспектной информации и превращение её в знания для эффективного управления бизнесом, осуществляемый конечными пользователями с помощью специальных технологий, методов и средств;

соответствующие технологии, дающие возможность организациям получать такие знания.

Бизнес-аналитика располагает различными методиками, а также средствами автоматизации, которые служат для поддержки принятия решений:

подсистемы информационно-поискового анализа;

подсистемы оперативной аналитической обработки данных (OLAP);

подсистемы так называемой «добычи данных», или интеллектуального анализа (методы и алгоритмы «Data Mining»).

Технологии OLTP и OLAP. Особенности, различия.

Традиционным подходом к организации баз данных и соответствующих обслуживающих приложений является OLTP-подход.

OLTP, или Online Transaction Processing, – это обработка транзакций в реальном времени. Структура такой базы данных сильно нормализована и оптимизирована для выполнения коротких идущих большим потоком транзакций, при этом клиенту требуется от системы минимальное время отклика. Обработываемый и сохраняемый OLTP-системой в течение дня объем данных может достигать не-скольких гигабайт. Примерами применения OLTP-подхода могут служить системы учета биржевых, банковских операций, системы бухгалтерского и складского учета и так далее.

Благодаря нормализации в таких системах значительно снижается избыточность данных и вычислительные потребности на операции обновления, что делает OLTP-системы идеальными для хранения данных. Однако сложность структуры таблиц и большие объемы накопленных данных приводят к снижению скорости выполнения сложных запросов на извлечение данных (например, посчитать прибыль организации по кварталам за последние пять лет), снижению производительности системы в целом. В результате эти системы оказываются непригодными для решения задач, диктуемых бизнес-аналитиками.

Поиск решения данной проблемы привел к формированию совершенно нового подхода, получившего название OLAP (Online Analytical Processing) – технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений. Цель таких систем – проверка гипотез пользователя-аналитика.

Как отмечалось ранее, OLTP-системы, приспособленные для хранения данных, оказались

непригодными для задач аналитиков. OLAP-системы же оптимизированы для выполнения операций чтения над большими объемами данных. Высокая скорость выполнения сложных аналитических запросов OLAP-системами связана с особенностями построения используемых ими многомерных структур (многомерные базы данных, или OLAP-кубы):

OLAP-системы строятся на базе денормализованных источников – хранилищ данных (принципы организации хранилищ данных рассмотрены в § 1.3); в итоге в базе данных OLAP могут содержаться избыточные данные, но в то же время положительным моментом в упрощении структуры связей таблиц является повышение скорости выполнения запросов;

многомерный куб содержит в себе не только сами данные, но и их агрегаты (обобщенные показатели) по различным измерениям (и иерархиям внутри измерений); то есть в базе данных OLAP хранятся заранее посчитанные системой показатели, которые потенциально могут потребоваться бизнес-аналитику.

Многомерность в OLAP-системах можно быть представить на трех уровнях:

многомерное представление данных – средства на стороне клиента, обеспечивающие многомерную визуализацию и манипулирование данными; слой многомерного представления абстрагирован от физической структуры данных и воспринимает данные как много- мерные;

многомерная обработка – средство (язык) формирования многомерных запросов (традиционный реляционный язык SQL здесь оказывается непригодным) и процессор, умеющий обработать и вы- полнить такой запрос;

многомерное хранение – средства физической организации данных, обеспечивающие эффективное выполнение многомерных запросов.

Первые два уровня в обязательном порядке присутствуют во всех OLAP-средствах. Третий уровень, хотя и является широко распространенным, не обязателен, так как данные для многомерного представления могут извлекаться и из обычных реляционных структур; процессор многомерных запросов в этом случае транслирует многомерные запросы в SQL-запросы, которые выполняются реляционной СУБД.

Виды OLAP-серверов.

В соответствии с требованием прозрачности OLAP-систем способ реализации многомерной модели скрыт от пользователя. Однако способ реализации важен, поскольку от него зависят производительность решения и требуемые ресурсы. Существует три основных способа реализации многомерной модели: MOLAP, ROLAP, HOLAP.

MOLAP (Multidimensional OLAP, или многомерный OLAP) – исходные и агрегатные данные хранятся в многомерной базе данных. Хранение данных в многомерных структурах позволяет манипулировать данными как многомерным упорядоченным массивом, благодаря чему скорость вычисления агрегатных значений одинакова для любого из измерений. Физически данные хранятся в «плоских» файлах, при этом куб представляется в виде одной плоской таблицы, в которую построено вписываются все комбинации элементов всех измерений с соответствующими им значениями мер.

В силу своих особенностей использование MOLAP является эффективным при следующих условиях:

объем исходных данных для анализа не слишком велик (не более нескольких гигабайт), то есть уровень агрегации данных достаточно высок;

набор информационных измерений стабилен (MOLAP чувствителен к изменению многомерной модели);

наименьшее время отклика системы на нерегламентированные запросы является критичным параметром (MOLAP обеспечивает высокую скорость поиска и выборки);

требуется использование сложных встроенных функций для выполнения вычислений над ячейками куба, возможность написания пользовательских функций (MOLAP легко

справляются с задачами включения в информационную модель разнообразных встроенных функций).

ROLAP (Relational OLAP, или реляционный OLAP) – исходные данные остаются в той же реляционной базе данных, где они изначально и находились. Агрегатные же данные помещают в специально созданные для их хранения служебные таблицы в той же базе данных.

Достоинства ROLAP:

в большинстве случаев корпоративные хранилища данных реализуются средствами реляционных СУБД, и инструменты ROLAP позволяют производить анализ непосредственно над ними;

в случае переменной размерности задачи, когда изменения в структуру измерений приходится вносить достаточно часто, ROLAP- системы с динамическим представлением размерности являются оптимальным решением, так как в них такие модификации не требуют физической реорганизации БД;

реляционные СУБД обеспечивают значительно более высокий уровень защиты данных и хорошие возможности разграничения прав доступа.

Недостатки ROLAP:

меньшая производительность в сравнении с MOLAP. Для обеспечения производительности, сравнимой с MOLAP, реляционные системы требуют тщательной проработки схемы базы данных и настройки индексов.

HOLAP (Hybrid OLAP, или гибридный OLAP) – исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные данные хранятся в многомерной базе данных. Серверы HOLAP применяют подход ROLAP для разреженных областей многомерного пространства и подход MOLAP – для плотных областей. Серверы HOLAP разделяют запрос на несколько под- запросов, направляют их к соответствующим фрагментам данных, комбинируют результаты, а затем предоставляют результат пользователю.

## Хранилища данных

### Понятие хранилища данных

Технология OLAP неразрывно связана с хранилищами данных. Билл Инмон, стоявший у истоков концепции хранилищ данных, в своей книге «Building the Data Warehouse» определяет хранилище данных как объектно-ориентированный, интегрированный, долгосрочный и отражающий изменения во времени набор данных, который предназначен для поддержки принятия управленческих решений.

Зачем строить отдельное хранилище для решения задач анализа и процессов принятия решений, когда вся нужная информация уже содержится в базах данных, внедренных на предприятии OLTP- систем? Ведь это решение приводит к формированию заведомо избыточной информации в информационной системе предприятия.

При ответе на этот вопрос необходимо охватить следующие аспекты:

анализировать данные OLTP-систем напрямую невозможно или затруднительно вследствие разрозненности разрозненностью данных, хранением их в форматах различных СУБД и в разных сегментах корпоративной сети; даже в случае, если на предприятии организовано централизованное хранение данных, аналитику довольно трудно разобраться в сложных, запутанных структурах БД, состоящих из десятков таблиц; сложные аналитические запросы к оперативной информации тормозят текущую работу компании, надолго блокируя таблицы и захватывая ресурсы сервера.

Приведенные факторы обуславливают необходимость организации на предприятии простой и понятной структуры, предоставляла бы «материал» для систем бизнес-анализа в одном месте. В качестве такой структуры и выступают хранилища данных.

Ральф Кимбалл, один из авторов концепции хранилищ данных, описывал хранилище

данных как «место, где люди могут получить доступ к своим данным». Он сформулировал основные требования к хранилищам данных:

поддержка высокой скорости получения данных из хранилища;

поддержка внутренней непротиворечивости данных;

возможность получения и сравнения так срезов данных;

наличие удобных утилит просмотра данных в хранилище;

полнота и достоверность хранимых данных;

поддержка качественного процесса пополнения данных.

В соответствии с классическим определением Инмона можно

выделить следующие свойства хранилищ данных:

объектная (предметная) ориентация – хранилища данных интегрируют информацию, отражающую различные точки зрения на предметную область;

интегрированность – на всем предприятии данные должны вводиться единообразно, либо требуется применять методы обеспечения согласованности информации на уровне хранилища данных;

хранение долгосрочного набора данных – обычно данные загружаются в хранилища данных с некоторым заданным интервалом, который может измеряться неделями или даже месяцами (в отличие от OLTP-систем);

отражение изменений во времени – все данные в хранилище соответствуют последовательным интервалам времени.

Чтобы понять место хранилищ данных в структуре систем поддержки принятия решений, достаточно рассмотреть схему, представленную на рис. 1.1. Хранилище данных является своеобразной прослойкой между транзакционными и оперативно-аналитическими системами. Данные, хранящиеся в системах OLTP, извлекаются, трансформируются и заносятся в хранилища данных, структура которых оптимизирована для задач бизнес-анализа (отчеты, OLAP-анализ, интеллектуальный анализ).

Примечание: весь процесс консолидации данных называется ETL (extraction, transformation, loading – извлечение, преобразование, загрузка). Для поддержки пользователей в процессе ETL компания Microsoft предоставляет средство SQL Server Integration Services (SSIS).

Киоск данных можно трактовать как мини-хранилище данных. Довольно часто киоски данных являются составными частями большого хранилища данных. Киоски данных представляют собой объектно-ориентированные склады данных, в которых хранятся отобранные в соответствии с определенной специализацией данные. Примером могут служить киоски данных с информацией о продажах, товарах на складе или, по сути, о любом предмете бизнес-деятельности, связанном с работой конкретного отдела. В противоположность этому хранилище данных функционирует на уровне предприятия и обычно обрабатывает данные всех отделов организации.

Хранилища данных организуют как на основе реляционной, так и многомерной модели данных.

Среди основных проблем, с которыми приходится сталкиваться при создании хранилищ данных, можно выделить следующие:

интеграция разнородных данных (данные поступают в хранилище из разнородных OLTP-систем, которые физически могут быть расположены в различных узлах сети);

эффективное хранение и обработка больших объемов данных (построение хранилища данных предполагает накопление данных за значительные периоды времени, что ведет к постоянному росту объемов дисковой памяти, а также росту объема оперативной памяти, требующейся для обработки этих данных);

организация многоуровневых справочников метаданных (конечным пользователям необходимы метаданные, описывающие структуру хранящихся данных, а также инструменты их визуализации);

обеспечение информационной безопасности (сводная информация о деятельности

компании, как правило, относится к коммерческой тайне и подлежит защите; также в хранилище данных могут содержаться персональные данные клиентов и сотрудников, которые также необходимо защищать; для выполнения этой функции должна быть разработана политика безопасности и программно-технические мероприятия по защите информации).

Подходы к организации хранилищ данных

Можно выделить следующие подходы к организации хранилищ данных:

физическое хранилище;

виртуальное хранилище.

В первом случае информация физически переносится из OLTP- системы в хранилище данных. В ходе этой загрузки данные фильтруются, поскольку не все из них имеют значение для проведения процедур анализа. В хранилище данных содержится обобщенная информация, которая в OLTP-системе отсутствует.

В системе с виртуальным хранилищем данные из OLTP-системы не копируются в единое хранилище, а извлекаются, преобразуются и интегрируются непосредственно при выполнении аналитических запросов в режиме реального времени. Фактически такие запросы напрямую передаются к OLTP-системе.

Достоинства виртуального хранилища:

минимизация объема хранимых данных;

работа с текущими, актуальными данными. Недостатки виртуального хранилища:

более высокое, по сравнению с физическим хранилищем данных время обработки запросов;

необходимость постоянной доступности всех OLTP- источников;

снижение быстродействия OLTP-систем;

OLTP-системы не ориентированы на хранение данных за длительный период времени, по мере необходимости данные выгружаются в архивные, поэтому не всегда имеется физическая возможность получения полного набора данных.

Проектирование хранилищ данных

Структура хранилища данных специфична в силу своего предназначения. Во-первых, как отмечалось ранее, схема хранилища данных денормализована. Во-вторых, хранилища данных строятся на базе двух типов таблиц – таблиц фактов и таблиц измерений.

Таблица фактов – основная таблица хранилища данных. Она содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться.

Встречающихся следующие основные типы фактов:

факты, основанные на отдельных событиях (на базе транзакций);

факты, связанные с «моментальными снимками» (основаны на состоянии объекта в определенные моменты времени; типичными примерами таких фактов являются объем продаж за день или дневная выручка);

факты, связанные с элементами документа (содержат подробную информацию об элементах определенного документа, например, количестве, цене, проценте скидки);

факты, связанные с событиями или состоянием объекта (события без подробностей о них).

Таблица фактов, как правило, содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений. Чаще всего это целочисленные значения либо значения типа «дата/время» –

ведь таблица фактов может содержать сотни тысяч или даже миллионы записей, и хранить в ней повторяющиеся текстовые описания, как правило, невыгодно - лучше поместить их в меньшие по объему таблицы измерений. При этом как ключевые, так и некоторые неключевые поля должны соответствовать будущим измерениям OLAP-куба. Помимо этого таблица фактов содержит одно или несколько числовых полей-фактов, на основании которых в дальнейшем будут получены агрегатные данные.

Для многомерного анализа пригодны таблицы фактов, содержащие как можно более

подробные данные (агрегатные данные все равно будут вычислены OLAP-средством). В таблице фактов нет никаких сведений о том, как группировать записи при вычислении агрегатных данных. Например, в ней есть идентификаторы продуктов или клиентов, но отсутствует информация о том, к какой категории относится данный продукт или в каком городе находится данный клиент. Эти сведения, в дальнейшем используемые для построения иерархий в измерениях куба, содержатся в таблицах измерений.

Таблицы измерений содержат неизменяемые или редко изменяемые данные. Эти таблицы содержат данные, называемые измерениями (dimensions), относительно которых будут анализироваться собранные в таблице фактов данные. Это может быть время (исследование временных тенденций), данные о клиентах (например, для анализа клиентского спроса и формирование для отдельных групп клиентов специальных предложений) и т.д. В подавляющем большинстве случаев эти данные представляют собой по одной записи для каждого члена нижнего уровня иерархии в измерении. Таблицы измерений также содержат как минимум одно описательное поле (обычно с именем члена измерения) и, как правило, целочисленное ключевое поле (обычно это суррогатный ключ) для однозначной идентификации члена измерения. Если будущее измерение, основанное на данной таблице измерений, содержит иерархию, то таблица измерений также может содержать поля, указывающие на «родителя» данного члена в этой иерархии.

Каждая таблица измерений должна находиться в отношении «один ко многим» с таблицей фактов.

Скорость роста таблиц измерений должна быть незначительной по сравнению со скоростью роста таблицы фактов.

Лабораторная работа № 2.

построение многомерного куба

Цель работы – получение навыков построения, редактирования структуры и просмотра данных многомерного куба средствами служб SQL Server Analysis Services.

## Основные теоретические сведения

### Терминология служб SQL Server Analysis Services

Службы SQL Server Analysis Services (SSAS) являются группой сервисов, которые используются для управления данными, хранящимися в хранилищах или киосках данных. Посредством служб SSAS данные из хранилища данных можно организовать в многомерные кубы с агрегированными данными, чтобы получить возможность создавать сложные запросы и подробные отчеты.

Основными особенностями служб SSAS являются следующие: удобство использования;

поддержка различных архитектур;

поддержка некоторых интерфейсов API.

Службы SSAS предоставляют мастеров почти для каждой задачи, выполняющейся в процессе проектирования и реализации хранилища данных. Например, мастер указания источников данных Data Source Wizard позволяет задать один или несколько источников данных, а мастер кубов Cube Wizard применяется для создания многомерных кубов и хранения агрегированных данных. Кроме этого, простота использования обеспечивается средством SQL Server Data Tools (SSDT), с помощью которого можно разрабатывать базы данных и другие объекты хранилищ данных. Этому способствует то, что средство SSDT предоставляет один общий интерфейс для разработки проектов SSAS, а также SQL Server Integration Services (SSIS) и SQL Server Reporting Services (SSRS).

В отличие от большинства других систем хранилищ данных, службы SSAS позволяют использовать ту архитектуру, которая наиболее отвечает имеющимся требованиям. В частности, можно выбирать между тремя архитектурами хранилищ данных (MOLAP, ROLAP и HOLAP).

Службы SSAS предоставляют много различных интерфейсов API для извлечения и доставки данных. Одним из этих интерфейсов является OLE DB для взаимодействия с OLAP, который позволяет обращаться к данным кубов SSAS.

Приведем наиболее важные термины служб SSAS: куб (cube), измерение (dimension), член (member), иерархия (hierarchy), ячейка (cell), уровень (level), группа мер (measure group), агрегация (aggregations), секция (partition).

Куб (cube) – это многомерная структура, содержащая все или часть данных из хранилища данных. Любой конкретный куб обычно имеет конкретный направленный предмет анализа. Хотя термин «куб» предполагает три измерения, многомерный куб обычно может иметь намного больше измерений. Каждый куб содержит все другие элементы, перечисленные выше.

Измерение (dimension) – это набор логически связанных атрибутов (хранящихся вместе в таблице измерения), которые подробно описывают меры (хранящиеся в таблице фактов). Например, Time, Product и Customer будут типичными измерениями, используемыми во многих приложениях бизнес-аналитики.

Примечание: одним из важных измерений куба является измерение Measures, которое содержит все меры, определенные в таблице фактов.

Каждое измерение содержит дискретные значения, называемые членами измерения (member). Например, членами измерения Product могут быть Computers, Disks и CPUs. Каждый член может быть вычисляемым, то есть его значение вычисляется в процессе выполнения, используя выражение, которое задается при определении члена. Поскольку вычисляемые члены не сохраняются на диске, то можно добавлять новые члены, не увеличивая объем соответствующего куба.

Иерархии (hierarchy) задают группирование множественных членов в каждом измерении. Они применяются для уточнения запросов, связанных с анализом данных.

Ячейки (cell) являются частями многомерного куба, которые определяются координатами (x, y и z для трехмерного куба). Это означает, что ячейка является набором, содержащим члены из каждого измерения.

Иерархии определяются на основе их уровней. Иными словами, уровни (level) описывают иерархию от самого высшего уровня данных (наиболее обобщенной) до самого низшего (наиболее подробной). В следующем списке приводятся возможные уровни иерархии для измерения Time: Quarter (Q1, Q2, Q3, Q4); Month (January, February, ...); Day (Day1, Day2, ...).

Меры (measure) – это числовые значения, такие как цена и количество, которые присутствуют в таблице фактов, но не используются в ее первичном ключе.

Группа мер (measure group) – это набор мер, которые совместно создают логическую единицу для целей деловой деятельности. Каждая группа мер создается на лету, используя соответствующую информацию из метаданных.

Конечной целью создания подобных кубов является минимизация времени обработки запросов, извлекающих требуемую информацию из фактических данных. Для реализации этой задачи кубы обычно содержат предварительно вычисленные итоговые данные, называемые агрегациями (aggregations). То есть куб охватывает пространство данных большее, чем фактическое – в нем существуют логические, вычисляемые точки. Вычислять значения точек в логическом пространстве на основе фактических значений позволяют функции агрегирования. Наиболее простыми функциями агрегирования являются SUM, MAX, MIN, COUNT.

Еще одной специфической чертой многомерных кубов является сложность определения точки начала координат. Решением этой проблемы является внедрение специального атрибута, объединяющего все элементы измерения. Этот атрибут (создается автоматически) содержит всего один элемент – All («Все»). Для простых функций агрегирования, например, суммы, элемент All эквивалентен сумме значений всех элементов фактического пространства данного измерения.

Для куба можно задать одну или несколько секций. Секции (partition) используются службами SSAS для управления и хранения данных и агрегированных данных для группы мер куба. Каждая группа мер имеет, по крайней мере, одну секцию, которая создается при определении этой группы мер. Секции являются мощным и гибким средством управления больших кубов.

Важной концепцией многомерной модели данных является под-пространство, или подкуб (sub cube). Подкуб представляет собой часть полного пространства куба в виде некоторой многомерной фигуры внутри куба. Так как многомерное пространство куба дискретно и ограничено, подкуб также дискретен и ограничен.

Разработка многомерного куба

Основным компонентом служб SSAS является средство управления SQL Server Data Tools (SSDT), которое предоставляет единую платформу для разработки различных приложений бизнес-аналитики. Средство SSDT основано на Visual Studio и поддерживает интегрированную платформу для разработчиков систем бизнес-аналитики.

С помощью средства SSDT можно не только создавать и управлять кубами, но также разрабатывать возможности для служб отчетов SQL Server Reporting Services (SSRS) и интеграционных служб SQL Server Integration Services (SSIS).

Примечание: пользовательский интерфейс средства SSDT подобен интерфейсу средства SQL Server Management Studio. Но эти два инструмента имеют разное назначение: средство SSDT применяется для разработки проектов бизнес-аналитики, тогда как основным назначением средства SQL Server Management Studio является работа с объектами баз данных в отношении бизнес-аналитики.

Процесс создания и обработки многомерного куба посредством средства SSDT состоит из следующих шагов:

Создание проекта бизнес-аналитики.

Определение источников данных.

Создание представлений источников данных.

Создание куба.

Проектирование агрегирования для хранилища.

Обработка куба.

Просмотр куба.

## **6. Учебно-методическое и информационное обеспечение дисциплины (модуля)**

### **6.1. Перечень учебной литературы, необходимой для освоения дисциплины (модуля)**

**Основная учебная литература:**

1. БАЗЫ ДАННЫХ -Санкт-Петербургский государственный электротехнический университет ЛЭТИ имени В.И. Ульянова (Ленина) (г. Санкт-Петербург) \\https://urait.ru/author-course/bazy-dannyh-514585(Дата обращения 25.05.2023)
2. Нестеров, С. А. Базы данных : учебник и практикум для среднего профессионального образования / С. А. Нестеров. — Москва : Издательство Юрайт, 2023. — 230 с. — (Профессиональное образование). — ISBN 978-5-534-11629-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/518507 (дата обращения: 25.05.2023).
3. Стружкин, Н. П. Базы данных: проектирование : учебник для вузов / Н. П. Стружкин, В. В. Годин. — Москва : Издательство Юрайт, 2023. — 477 с. — (Высшее образование). — ISBN 978-5-534-00229-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511019 (дата обращения: 25.05.2023).

### Дополнительная учебная литература:

1. Стружкин, Н. П. Базы данных: проектирование. Практикум : учебное пособие для вузов / Н. П. Стружкин, В. В. Годин. — Москва : Издательство Юрайт, 2023. — 291 с. — (Высшее образование). — ISBN 978-5-534-00739-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/512160> (дата обращения: 25.05.2023).
2. Стасышин, В. М. Базы данных: технологии доступа : учебное пособие для среднего профессионального образования / В. М. Стасышин, Т. Л. Стасышина. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 164 с. — (Профессиональное образование). — ISBN 978-5-534-09888-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/516927> (дата обращения: 25.05.2023).

### 6.2. Перечень электронных библиотечных систем, современных профессиональных баз данных и информационных справочных систем

№ п/п	Наименование документа с указанием реквизитов
1	Договор на доступ к ЭБС ZNANIUM.COM между БашГУ в лице директора СФ БашГУ и ООО «Знаниум» № 3/22-эбс от 05.07.2022
2	Договор на доступ к ЭБС «ЭБС ЮРАЙТ» (полная коллекция) между БашГУ в лице директора СФ БашГУ и ООО «Электронное издательство ЮРАЙТ» № 1/22-эбс от 04.03.2022
3	Договор на доступ к ЭБС «Университетская библиотека онлайн» между БашГУ и «Нексмедиа» № 223-950 от 05.09.2022
4	Договор на доступ к ЭБС «Лань» между БашГУ и издательством «Лань» № 223-948 от 05.09.2022
5	Договор на доступ к ЭБС «Лань» между БашГУ и издательством «Лань» № 223-949 от 05.09.2022
6	Соглашение о сотрудничестве между БашГУ и издательством «Лань» № 5 от 05.09.2022
7	ЭБС «ЭБ БашГУ», бессрочный договор между БашГУ и ООО «Открытые библиотечные системы» № 095 от 01.09.2014 г.
8	Договор на БД диссертаций между БашГУ и РГБ № 223-796 от 27.07.2022
9	Договор о подключении к НЭБ и о предоставлении доступа к объектам НЭБ между БашГУ в лице директора СФ БашГУ с ФГБУ «РГБ» № 101/НЭБ/1438-П от 11.06.2019
10	Договор на доступ к ЭБС «ЭБС ЮРАЙТ» (полная коллекция) между УУНиТ в лице директора СФ УУНиТ и ООО «Электронное издательство ЮРАЙТ» № 1/23-эбс от 03.03.2023

### Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»)

№ п/п	Адрес (URL)	Описание страницы
1	<a href="http://wikipedia.org">http://wikipedia.org</a>	Полноценный сайт по всей нужной информации.
2	<a href="http://intuit.ru">intuit.ru</a>	Бесплатное дистанционное обучение в национальном открытом институте "Интуит".

### 6.3. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

Наименование программного обеспечения
Anaconda Интегрированная среда разработки ПО. Свободно-распространяемое ПО
Visual Studio Community Интегрированная среда разработки ПО. Свободно-распространяемое ПО
Браузеры Google Chrome, Mozilla Firefox. Свободно-распространяемое ПО
Файловый архиватор 7z. Свободно-распространяемое ПО
Язык программирования Python и среда разработки Jupiter Notebook (вместе с библиотеками numpy, scikit-learn, pandas)
Программное обеспечение для создания и просмотра pdf-документов Adobe Reader
Программное обеспечение для подготовки слайдов лекций MS PowerPoint, MS Word
PyCharm Community Интегрированная среда разработки ПО. Свободно-распространяемое ПО
Операционная система Windows

### 7. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Тип учебной аудитории	Оснащенность учебной аудитории
Учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория текущего контроля и промежуточной аттестации, учебная аудитория групповых и индивидуальных консультаций	Доска, учебная мебель, проектор, экран, учебно-наглядные пособия.
Учебная аудитория для проведения занятий семинарского типа, учебная аудитория текущего контроля и промежуточной аттестации, учебная аудитория групповых и индивидуальных консультаций	Доска, проектор, экран, учебная мебель, компьютеры, учебно-наглядные пособия.
Учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория текущего контроля и промежуточной аттестации, учебная аудитория групповых и индивидуальных консультаций, учебная аудитория курсового проектирования (выполнения курсовых работ)	Доска, учебная мебель, проектор, экран, компьютеры, учебно-наглядные пособия.