

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

ХРАНИЛИЩА ДАННЫХ

Методические указания к лабораторным работам
и организации самостоятельной работы для студентов направления
«Бизнес-информатика»
(уровень бакалавриата)

2018

Жуковский Олег Игоревич

Хранилища данных: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направления «бизнес-информатика» (уровень бакалавриата) / О.И. Жуковский. – Томск, 2018. – 25 с.

© Томский государственный университет систем
управления и радиоэлектроники, 2018
© Жуковский О.И., 2018

Оглавление

1 Введение	4
2 Методические указания к проведению лабораторных работ 5	
2.1 Лабораторная работа «Формулирование требований»	5
2.2 Лабораторная работа «Моделирование данных»	7
2.3 Лабораторная работа «Определение процедур извлечения, преобразования и загрузки данных»	9
2.4 Лабораторная работа «Проектирование аналитических отчетов»	11
2.5 Лабораторная работа «Сопровождение и модификация хранилища данных»	13
2.6 Лабораторная работа «Построение функциональной модели процесса разработки хранилища данных».....	15
3 Методические указания для организации самостоятельной работы	17
3.1 Общие положения.....	17
3.2 Проработка лекционного материала.....	17
3.3 Изучение тем (вопросов) теоретической части дисциплины, вынесенных для самостоятельной подготовки	18
3.3.1 Тиражирование данных.....	18
3.3.2 Витрина данных	18
3.3.3 Федеративное хранилище данных	19
3.3.4 Факторы, влияющие на структуру проекта создания хранилища данных.	19
3.3.5 Связь хранилищ данных и OLAP-систем.....	20
3.3.6 Особенности задачи кластеризации.....	21
3.4 Подготовка к лабораторным работам.....	21
3.5 Выполнение индивидуального задания.....	22
3.6 Подготовка к зачету.....	25
3.7 Подготовка к экзамену	25
4 Рекомендуемая литература	25

1 Введение

Выполнение лабораторных работ и самостоятельная работа направлены на овладение знаниями современных технологий и методов анализа и проектирования процесса разработки автоматизированных информационных систем, ориентированных на анализ данных студентами направления подготовки бакалавров «Бизнес-информатика».

В результате изучения дисциплины студент должен:

Знать: основные принципы построения систем, ориентированных на анализ данных; базовые понятия современных технологий интеллектуального анализа данных; модели данных, используемые для построения хранилищ; особенности построения систем на основе хранилищ данных; основные принципы OLAP, назначение и область применения методов интеллектуального анализа данных; основные этапы процесса создания хранилищ данных.

Уметь: проводить анализ цикла построения информационных систем на основе концепции хранилищ данных; определять необходимость применения технологий хранилищ данных; выбирать структуры хранилищ данных;

Владеть: методами и программными средствами разработки функциональных моделей процессов проектирования и создания хранилищ данных автоматизируемой организации.

Каждой лабораторной работе предшествует самостоятельная работа, выполняемая студентом перед аудиторным занятием. Содержание самостоятельной работы по подготовке к каждой лабораторной работе приводится в данном учебно-методическом пособии. В значительной мере самостоятельная работа состоит в сборе и анализе информации об объекте управления, для анализа состояния и процессов которого с помощью системы поддержки принятия решений на основе хранилища данных и будут использоваться рассматриваемые в данном курсе информационные технологии.

Для сбора информации можно использовать поиск информации в литературных источниках и Интернете. Форма контроля выполнения лабораторной работы: демонстрация преподавателю полученных результатов, собеседование, ответы на вопросы, выполнение дополнительных заданий.

2 Методические указания к проведению лабораторных работ

2.1 Лабораторная работа «Формулирование требований»

Цель работы

Научиться планировать процесс формулирования требований при разработке хранилища данных.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

На проверку должна быть представлена функциональная модель, показывающая особенности этапа формулирования требований процесса разработки хранилища данных в предметной области, соответствующей индивидуальному заданию. Также необходимо представить в письменном виде сами сформулированные требования.

Теоретические основы

Главной задачей данного этапа процесса разработки хранилища данных (ХД) является идентификация требований заказчика ХД и оформление их в виде документа «Каталог требований». Обычно сбор требований осуществляется путем опроса групп потенциальных пользователей ХД на специальных совещаниях и переговорах. Конечные пользователи, как правило, не знакомы с концепцией ХД и процессом складирования данных. Поэтому для успешного решения этой задачи важна помощь лиц, принимающих решения (ЛПР), т.е. руководителей организации. На этом этапе определяются:

- масштаб проекта создания ХД (границы предметной области);
- перечень и содержание отчетов;
- требования по анализу данных (перечень задач анализа данных);
- требования к аппаратному обеспечению;
- требования к системному программному обеспечению;
- значения базовых параметров Хранилища данных (скорость обработки запросов, объемы используемых данных, актуализация данных, производительность системы и т.д.);
- требования к квалификации персонала (программа обучения персонала);
- перечень источников данных;

- конкретизация плана реализации проекта разработки ХД (определяется дата завершения проекта).

В дополнение, основываясь на собранной выше информации, составляется план восстановления хранилища данных в случае аварийных сбоях. Разрабатывается стратегия архивирования и восстановления ХД.

Временные требования. Время выполнения этапа — от двух недель до двух месяцев.

Результатом выполнения этапа являются каталог требований, утвержденный заказчиком, и уточненный план проекта, который точно определяет используемые ресурсы и даты контрольных точек проверки хода выполнения проекта.

Потенциальные опасности. Этап формулирования требований часто оказывается одним из самых узких мест проекта создания ХД. Причина состоит в конфликте внутрикорпоративных интересов и в необходимости наладить коммуникации для успешного выполнения и этапа и проекта в целом.

По определению складирование данных предполагает интеграцию в ХД данных из нескольких источников (подразделений организации). Позиция и взгляды подразделений на бизнес-информацию (производство данных, их верификацию, поставку данных в ХД, разграничение доступа к данным и т.д.) могут сильно расходиться и даже быть диаметрально противоположными. Даже если команда разработчиков предложит блестящее решение по созданию ХД, его реализация может споткнуться о нежелание определенных групп потенциальных пользователей предоставлять данные в ХД или конструктивно участвовать в определении требований к системе складирования данных.

Если не удастся наладить коммуникации между участниками процесса, то все усилия по созданию ХД будут потрачены впустую и проект никогда не будет завершен в установленные сроки, а в худшем случае просто будет провален.

Одним из способов избежать такой ловушки является непосредственное вовлечение руководителей организации в процесс реализации проекта. Следует заручиться поддержкой влиятельного покровителя из числа высшего руководства, чтобы можно было влиять на позицию подразделений сотрудничать с командой разработчиков ХД.

Для успешного выполнения лабораторной работы необходимо изучить разделы 4.1 и 4.2 методического пособия [1].

Порядок выполнения лабораторной работы

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных этапов для реализации небольшого или среднего по масштабу проекта. Акцентируйте внимание на роли этапа Формулирования требований, связи его с другими этапами.

Повторите материал раздела 2 методических указаний [3], посвященный построению функциональных моделей и создайте функциональную модель, представляющую процесс Формулирования требований в вашей предметной области. Например, какие данные и из каких источников вашей предметной области будут извлекаться, какие аналитические отчеты будут создаваться. Декомпозируйте не менее трех блоков диаграммы проектного цикла. Не забывайте о цели модели и точке зрения автора модели.

Полученную функциональную IDEF0-модель продемонстрируйте преподавателю.

2.2 Лабораторная работа «Моделирование данных»

Цель работы

Научиться создавать концептуальную модели данных для проектируемого ХД.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Результатом выполнения этапа являются логическая и физическая модели данных.

На проверку должен быть представлен перечень источников данных и их описание, а также концептуальная модель данных проектируемого хранилища данных. Модель должна содержать все необходимые данные, используемые для построения аналитических отчетов по итогам анализа данных из хранилища.

Теоретические основы

Главной задачей этапа является разработка логической и физической моделей данных для ХД. Этот этап — один из самых важных для проекта создания ХД. Ошибки и просчеты, допущенные на этом этапе, будет очень сложно исправить на последующих этапах. Кроме того, подобные просчеты могут привести к пересмотру всего проекта и, следовательно, к его фактическому провалу.

При выполнении этого этапа сначала строится логическая модель данных, которая впоследствии преобразуется в физическую модель.

Одной из подзадач этого этапа являются идентификация и описание источников данных, которые также могут стать подзадачей этапа определения процедур извлечения, преобразования и загрузки данных в ХД (ETL-процессов).

Временные требования. Время выполнения этого этапа — от двух недель до двух месяцев.

Результатом выполнения этапа являются перечень источников данных и их описание, а также логическая и физическая модели данных.

Потенциальные опасности. Самой большой опасностью при выполнении этого этапа является самоуверенность проектировщиков ХД. Во многих случаях даже опытные проектировщики допускают от двух до пяти ошибок в структуре данных (потерь существенных семантических зависимостей в данных) на проект. Причиной таких ошибок часто становится недостаточная осведомленность проектировщиков о предметной области ХД и низкое качество информации, поставляемой аналитиками предметной области.

Бизнес-аналитики предметной области могут не предоставить полной информации о функциональных зависимостях в данных, что приведет в результате к проектированию частично неправильной схемы данных. В этом случае схема правится «на лету» на последующих этапах выполнения проекта, что может привести к пересмотру архитектуры приложений и процессов ETL. Иногда ошибки носят принципиальный характер и приводят к полному пересмотру фрагмента схемы данных. Например, как в случае, когда пропущена зависимость вложения на подмножествах в данных (типа «часть-целое»).

Хорошей предупредительной мерой для предотвращения подобных ситуаций является привлечение квалифицированных экспертов, особенно в случае, если проект разработки ХД выполняется силами самой организации.

Для успешного выполнения лабораторной работы необходимо повторить раздел 2.6 учебно-методического пособия по базам данных [2] и изучить разделы 2.2 – 2.4 методического пособия [1].

Порядок выполнения лабораторной работы

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных этапов для реализации небольшого или среднего по масштабу проекта. Акцентируйте внимание на роли этапа Моделирование данных, связи его с другими этапами.

Определите тип и содержание данных, которые будут поступать из источников для хранилища вашей предметной области, согласно индивидуальному заданию. Согласуйте с преподавателем содержание и особенности представления анализируемых данных в хранилище. Опишите необходимые для ХД сущности и их атрибуты. Обратите внимание на покрытие планируемых аналитических отчетов набором данных, описываемых в создаваемой модели. Создайте концептуальную модель данных проектируемого ХД.

Представьте полученную концептуальную модель данных и описание ее атрибутов преподавателю.

2.3 Лабораторная работа «Определение процедур извлечения, преобразования и загрузки данных»

Цель работы

Научиться планировать процесс создания процедур извлечения, преобразования и загрузки данных.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

На проверку должна быть представлена функциональная модель процесса создания процедур извлечения, преобразования и загрузки данных, а также состав и описание возможных процедур извлечения, преобразования и загрузки данных проектируемого вами ХД согласно особенностям выбранной предметной области.

Теоретические основы

Главной задачей этапа является идентификация и определение процедур извлечения, очистки (фильтрации), преобразования и загрузки данных. Это весьма трудоемкий этап, не столько по временным затратам, сколько по усилиям, предпринимаемым по анализу структур данных источников и подающих систем.

Исключительно редким является случай, когда ХД данных создается на голом месте, т.е. в подразделениях отсутствуют автоматизированные подсистемы обработки данных. Как правило, данные уже существуют (в том или ином виде). Их нужно собрать, согласовать, привести к единому формату, агрегировать и загрузить в ХД. По этой причине этот этап является характерным для проекта создания ХД. Также следует помнить о том, что сам процесс подготовки и загрузки данных в создаваемое ХД может занять более половины времени, отведенного на реализацию проекта, особенно в проектах большого масштаба.

Временные требования. Время выполнения этапа — от одной недели до полутора месяцев.

Результатом выполнения этапа являются схема соответствия данных подающих систем и ХД, программы или ETL-инструменты.

Потенциальные опасности. Первой потенциальной опасностью при выполнении этого этапа является недооценка временных параметров. Обычно на выполнение этапа выделяют немного времени. Процедуры извлечения, преобразования и загрузки данных в ХД оказывают непосредственное влияние на качество и полноту предоставляемой информации конечным пользователям. Недостаточное внимание к разработке этих процедур может вызвать негативные реакции у пользователей после получения ими отчетов.

Второй потенциальной опасностью является стремление команды разработчиков сделать процесс ETL как можно более всеобъемлющим, мотивируя свои действия стремлением обеспечить качество данных. Следует помнить, что главная цель ETL-процесса — оптимизации скорости загрузки данных в ХД.

Для успешного выполнения лабораторной работы необходимо изучить разделы 2.2 – 2.4 и 3.2– 3.4 базового методического пособия [1].

Порядок выполнения лабораторной работы

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных эта-

пов для реализации небольшого или среднего по масштабу проекта. Акцентируйте внимание на роли этапа, связанного с определением процедур извлечения и очистки данных, связи его с другими этапами.

Проведите анализ данных, поступающих из первичных источников. Обратите внимание на возможные дефекты данных – ошибки уровня ячейки, уровня записи, таблицы и базы данных

Опишите содержание возможных процедур очистки. Проведите поиск в сети интернет готовых программных решений для планируемых вами процедур очистки. При отсутствии подходящих решений сформулируйте требования к программам очистки, которые необходимо заново создать.

Повторите материал раздела 2 методических указаний [3], посвященный построению функциональных моделей и создайте функциональную модель, представляющую процесс определения процедур извлечения, преобразования и загрузки данных при создании ХД в вашей предметной области. Например, какие данные и из каких источников вашей предметной области будут извлекаться, преобразовываться и загружаться в хранилище. Декомпозируйте не менее трех блоков диаграммы проектного цикла. Не забывайте о цели модели и точке зрения автора модели.

Полученную функциональную IDEF0-модель продемонстрируйте преподавателю.

2.4 Лабораторная работа «Проектирование аналитических отчетов»

Цель работы

Научиться моделировать процесс проектирования аналитических отчетов создаваемого хранилища данных.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

На проверку должна быть представлена функциональная модель процесса проектирования аналитических отчетов создаваемого хранилища данных, согласованная с вашим индивидуальным заданием.

Теоретические основы

Главной задачей выполнения этого этапа является проектирование и разработка аналитических отчетов на спроектированной структуре данных. Это также специфичный этап для проекта ХД.

Перечень требуемых отчетов содержится в «Каталоге требований», и их разработка решает в целом поставленную задачу. Однако следует помнить, что потенциальные пользователи не всегда точно знают, что они хотят увидеть в отчетах. С другой стороны, как правило, собранные данные предоставляют большие возможности по формированию разнообразных отчетов, чем это зафиксировано в «Каталоге требований». Здесь должен быть найден разумный компромисс.

Временные требования. Время выполнения этого этапа зависит от числа разрабатываемых отчетов. В зависимости от сложности отчета его разработка занимает от 4 часов до двух недель.

Результатом выполнения этапа станут спецификация кубов данных (измерения и метрики) и разработанные отчеты.

Потенциальные опасности. Потенциальной опасностью при выполнении этого этапа является то, что не уделяется достаточного внимания оптимизации времени получения отчета. Конечный пользователь не любит долго ждать. А если он получал такой отчет на своей настольной системе быстрее, то мнение о ХД у него будет, мягко говоря, неадекватное.

Хороший способ избежать такой опасности — тестирование разработанных отчетов с целью минимизации времени их получения.

Для успешного выполнения лабораторной работы необходимо изучить разделы 5.1 и 5.4 методического пособия [1].

Порядок выполнения лабораторной работы

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных этапов для реализации небольшого или среднего по масштабу проекта. Акцентируйте внимание на роли этапа планирования аналитических отчетов, связи его с другими этапами.

Выберите показатели предметной области вашего индивидуального задания, которые могут играть роль измерений и параметры, которые могут играть роль анализируемых мер. Согласуйте конечный набор с преподавателем.

Решите, как будут представлены данные в проектируемом ХД – в виде гиперкуба или поликуба.

Повторите материал раздела 2 методических указаний [3], посвященный построению функциональных моделей и создайте функциональную модель, представляющую процесс проектирования аналитических отчетов при создании ХД в вашей предметной области. Например, какие данные и из каких источников вашей предметной области будут использоваться для

определения мер, в какие отчеты они будут входить. Декомпозируйте не менее трех блоков диаграммы проектного цикла. Не забывайте о цели модели и точке зрения автора модели.

Полученную функциональную IDEF0-модель продемонстрируйте преподавателю.

2.5 Лабораторная работа «Сопровождение и модификация хранилища данных»

Цель работы

Получить навыки моделирования процесса сопровождения и модификации хранилища данных.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

На проверку должна быть представлена функциональная модель процесса сопровождения и модификации создаваемого хранилища данных, согласованная с вашим индивидуальным заданием.

Теоретические основы

Обычно в проектный цикл включают еще два этапа — этап сопровождения ХД и этап его модификации. Это важные этапы в жизни ХД. Однако, как показывает опыт, целесообразно выделять их в самостоятельные проекты.

Было бы большой ошибкой поручать разработчикам ХД его сопровождение. Процессы сопровождения ХД требуют от ИТ-специалистов иной квалификации, чем процессы его разработки. Чтобы сопровождать, не обязательно уметь писать программы, но обязательно нужно уметь их настраивать и использовать.

Если необходимость в модернизации ХД возникает спустя несколько месяцев после сдачи его в эксплуатацию, это говорит о том, что проект не был успешным. Потребность в модернизации реально может сформироваться спустя шесть месяцев после интенсивной его эксплуатации, когда проверены его возможности, прочувствована отдача и видны новые перспективы использования, т.е. когда ХД стало частью производственного процесса. При этом не факт, что та же команда разработчиков будет заниматься его модернизацией.

Процесс разработки ХД может быть представлен в виде модели бизнес-процессов. Бизнес-модель процесса разработки позволяет:

отобразить субъективное мнение руководителя ИТ и некоторых участников команды на процесс проектирования конкретного ХД;

- учесть особенности ИТ-проекта, в рамках которого проектируется ХД;
- достаточно быстро составить план проектирования конкретного ХД;
- просчитать длительность проектных работ (создать временную модель проектирования).

Для успешного выполнения лабораторной работы необходимо изучить разделы 4.1 и 4.2 методического пособия [1].

Порядок выполнения лабораторной работы

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных этапов для реализации небольшого или среднего по масштабу проекта. Акцентируйте внимание на роли этапа сопровождения и модификации хранилища данных, связи его с другими этапами.

Изучите возможные изменения источников данных вашего ХД и составьте примерный план проверок их состояния для возможной инициации процессов модификации разрабатываемого ХД.

Особое внимание уделите источникам, формирующим измерения многомерной модели хранилища, задействованные в формировании аналитических отчетов. Сделайте наброски предполагаемого плана сопровождения и модификации.

Повторите материал раздела 2 методических указаний [3], посвященный построению функциональных моделей и создайте функциональную модель, представляющую процесс сопровождения и модификации ХД для предметной области в соответствии с индивидуальным заданием. Например, какие источники и с какой периодичностью будут проверяться на наличие изменений в структуре загружаемых данных и каких в первую очередь. Декомпозируйте не менее трех блоков диаграммы проектного цикла. Не забывайте о цели модели и точке зрения автора модели.

Полученную функциональную IDEF0-модель продемонстрируйте преподавателю.

2.6 Лабораторная работа «Построение функциональной модели процесса разработки хранилища данных»

Цель работы

Научиться строить функциональные модели процессов разработки сложных информационных систем, к числу которых относятся системы на основе технологии хранилищ данных, включаемых в цепочку цифровых технологий принятия управленческих решений. Подобные модели позволяют провести анализ функций предполагаемой системы и основных информационных объектов, поддерживаемых системой в случае ее создания.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

На проверку должен быть представлена функциональная модель, показывающая основные этапы процесса разработки хранилища данных в предметной области, соответствующей индивидуальному заданию.

Теоретические основы

Для успешного выполнения лабораторной работы необходимо изучить главы 3 и 4 учебного пособия [1] и раздел 2 методических указаний [3], посвященный построению функциональных моделей.

Порядок выполнения лабораторной работы

Процесс разработки ХД может быть представлен в виде модели бизнес-процессов. Бизнес-модель процесса разработки позволяет:

отобразить субъективное мнение руководителя ИТ и некоторых участников команды на процесс проектирования конкретного ХД;

- учесть особенности ИТ-проекта, в рамках которого проектируется ХД;
- достаточно быстро составить план проектирования конкретного ХД;
- просчитать длительность проектных работ (создать временную модель проектирования).

Рассмотрите типовую модель бизнес-процессов разработки ХД, которая может быть положена в основу реализации любого конкретного проекта. Эта модель содержит минимально достаточное число обязательных этапов для реализации небольшого или среднего по масштабу проекта.

В проектный цикл разработки ХД обычно включаются следующие типовые процессы (этапы):

- формулирование требований;
- создание вычислительной среды;
- моделирование данных;
- определение процедур извлечения, преобразования и загрузки данных;
- проектирование аналитических отчетов;
- разработка приложений ХД;
- настройка производительности;
- проверка качества;
- передача системы складирования данных в эксплуатацию

Повторите материал раздела 2 методических указаний [3], посвященный построению функциональных моделей и создайте функциональную модель, представляющую процесс в целом, декомпозируйте те блоки, которые представляют действия, требующие описания менеджера бизнес проекта, знакомого с информационными технологиями. Например, какие данные и из каких источников вашей предметной области будут извлекаться, преобразовываться и загружаться в хранилище. Декомпозируйте не менее трех блоков диаграммы проектного цикла. Не забывайте о цели модели и точке зрения автора модели.

Создаваемая функциональная модель должна содержать ключевые моменты процесса проектирования ХД, разработанные в лабораторных работах, посвященных отдельным этапам процесса проектирования ХД.

Полученную функциональную IDEF0-модель продемонстрируйте преподавателю.

3 Методические указания для организации самостоятельной работы

3.1 Общие положения

Целями самостоятельной работы является систематизация, расширение и закрепление теоретических знаний, приобретение навыков научно-исследовательской и производственно-технологической деятельности.

Самостоятельная работа по дисциплине «Хранилища данных» включает следующие виды активности студента:

- проработка лекционного материала;
- изучение тем (вопросов) теоретической части дисциплины, вынесенных для самостоятельной подготовки;
- подготовка к лабораторным работам;
- выполнение индивидуального задания;
- подготовка к зачету;
- подготовка к экзамену.

3.2 Проработка лекционного материала

Для проработки лекционного материала студентам рекомендуется воспользоваться конспектом, сопоставить записи конспекта с соответствующими разделами методического пособия [1].

Целесообразно ознакомиться с информацией, представленной в файлах, содержащих презентации лекций, предоставляемых преподавателем.

Для проработки лекционного материала студентам, помимо конспектов лекций, рекомендуются следующие главы учебно-методического пособия [1] по разделам курса:

- системы поддержки принятия решений – глава 1;
- концепция хранилища данных – глава 2;
- архитектура хранилищ данных – глава 3;
- разработка и проектирование хранилищ данных – глава 4;
- OLAP-системы – глава 5;
- интеллектуальный анализ данных – глава 6.

При изучении учебно-методического пособия [1] студенту рекомендуется самостоятельно ответить на вопросы, приводимые в конце каждой главы. Рекомендуется сформулировать вопросы преподавателю и задать их

либо посредством электронной образовательной среды вуза, либо перед началом следующей лекции.

3.3 Изучение тем (вопросов) теоретической части дисциплины, вынесенных для самостоятельной подготовки

3.3.1 Тиражирование данных

Перечень вопросов, подлежащих изучению

- Особенности распределенной обработки данных;
- Процесс репликации данных как центральное звено тиражирования.
- Преимущества и недостатки процесса тиражирования данных.

Методические рекомендации по изучению

В первую очередь обратите внимание на отказ от распределенности данных и обязательном наличии в процессе использования данных регламентных требований, позволяющих в определенный промежуток времени работать на конкретных копиях базы данных как на основном экземпляре базы данных. Подумайте, в каких сферах деятельности присутствуют процессы, эффективная информационная поддержка которых допускает использование технологии «тиражирования данных».

Рекомендуемые источники

Для подготовки к изучению особенностей «тиражирования данных» необходимо ознакомиться с материалом, изложенным в главе 1 учебного пособия [1]. Кроме того, рекомендуется познакомиться с трактовками данного понятия, представленными в сети Интернет.

3.3.2 Витрина данных

Перечень вопросов, подлежащих изучению

- Отличия витрин данных от полномасштабных хранилищ;
- Основные преимущества подхода витрин данных в создании систем поддержки принятия решений организации;
- Целесообразность использования Витрин данных в связке с Хранилищем данных.

Методические рекомендации по изучению

Обратите внимание, что самостоятельные Витрины данных часто появляются в организации исторически и встречаются в крупных организациях с большим количеством независимых подразделений, решающих собственные аналитические задачи

Рекомендуемые источники

Для подготовки к изучению особенностей обработки информации необходимо ознакомиться с материалом, изложенным в разделе 2.1 учебного пособия [1].

3.3.3 Федеративное хранилище данных

Перечень вопросов, подлежащих изучению

- особенности архитектуры федеративного хранилища данных;
- состав экземпляра федеративного хранилища;
- масштаб систем, позволяющих эффективно использовать данный подход.

Методические рекомендации по изучению

Обратите внимание, что каждый из экземпляров федеративного ХД хранит копию базовой бизнес-модели и общие основные данные (common master dat), причем каждое ХД более высокого уровня содержит итоговые транзакционные данные более низкого уровня.

Рекомендуемые источники

Для подготовки к изучению особенностей «федерации» хранилищ данных необходимо ознакомиться с материалом, изложенным в разделе 3.5 учебного пособия [1].

3.3.4 Факторы, влияющие на структуру проекта создания хранилища данных

Перечень вопросов, подлежащих изучению

- Необходимость регулярного пересмотра функций и реинжиниринга OLTP-систем организации;
- Продолжительность периода использования ХД;
- Вес разработки каждого структурного компонента в проекте ХД.

Методические рекомендации по изучению

В первую очередь обратите внимание на цели процесса планирования создания ХД, поскольку ХД предназначено жить, даже если половина бизнес направлений организации будет закрыта, а на их смену придут новые. Информация не теряется, она остается и будет доступна для анализа предыдущего опыта организации. Весьма расточительно планировать ХД так, чтобы через два года его переделывать.

Рекомендуемые источники

Для подготовки к изучению видов разметки необходимо ознакомиться с материалом, изложенным в разделе 4.2 учебного пособия [1].

3.3.5 Связь хранилищ данных и OLAP-систем

Перечень вопросов, подлежащих изучению

- особенности организации концептуального многомерного представления данных;
- особенности архитектуры хранилищ данных для многомерного анализа данных;
- достоинства и недостатки многомерных баз данных.

Методические рекомендации по изучению

В первую очередь обратите внимание, что в концепции хранилища данных нет постановки вопросов, связанных с организацией эффективного анализа данных и предоставления доступа к ним. Эти задачи решаются подсистемами анализа. Обратите внимание, какой способ работы с данными наиболее подходит аналитику, пользователю системы поддержки принятия решений на базе хранилища данных. Сравните преимущества и недостатки многомерного и реляционного подхода в организации баз данных хранилища данных.

Рекомендуемые источники

Для подготовки к изучению особенностей использования многомерных хранилищ данных в системах поддержки принятия решения необходимо ознакомиться с материалом, изложенным в разделе 5.5 учебного пособия [1]. Также, в рамках изучения современных систем обработки данных, ориентированных на использование технологии хранилищ данных, рекомендуется познакомиться с материалами сайта компании "Интерфейс", одного из ведущих российских поставщиков инструментальных средств и решений для создания корпоративных информационных систем, разработки прило-

жений, управления проектами, реинжиниринга деятельности предприятий, OLAP – <http://www.interface.ru>

3.3.6 Особенности задачи кластеризации

Перечень вопросов, подлежащих изучению

- Цель моделирования данных;
- Области применения кластеризации;
- Понятия, характерные для задач кластеризации.

Методические рекомендации по изучению

В первую очередь обратите внимание, что задача кластеризации заключается в поиске независимых групп (кластеров) и их характеристик во всем множестве анализируемых данных. Решение этой задачи помогает лучше понять данные. Кроме того, группировка однородных объектов позволяет сократить их число, а следовательно, и облегчить анализ.

Рекомендуемые источники

Для подготовки к изучению особенностей задачи кластеризации необходимо ознакомиться с материалом, изложенным в разделе 6.2 учебного пособия [1].

3.4 Подготовка к лабораторным работам

Для подготовки к лабораторной работе «Формулирование требований» студенту необходимо:

- изучить разделы 4.1 и 4.2 методического пособия [1];
- согласовать с преподавателем набор и состав источников данных, соответствующих автоматизируемому процессу предметной области индивидуального задания.

Для подготовки к лабораторной работе «Моделирование данных» студенту необходимо:

- повторите раздел 2.6 учебно-методического пособия по базам данных [2] и изучите разделы 2.2 – 2.4 методического пособия [1];
- определить тип и содержание данных ХД предметной области, согласно индивидуальному заданию.
- согласовать с преподавателем содержание и особенности представления анализируемых данных в хранилище.

Для подготовки к лабораторной работе «Определение процедур извлечения, преобразования и загрузки данных» студенту необходимо:

- изучить разделы 2.2 – 2.4 и 3.2– 3.4 методического пособия [1];
- провести анализ данных, поступающих из первичных источников;
- описать содержание возможных процедур очистки.

Для подготовки к лабораторной работе «Проектирование аналитических отчетов» студенту необходимо:

- изучить разделы 5.1 и 5.4 методического пособия [1].
- выбрать и согласовать с преподавателем показатели предметной области вашего индивидуального задания, которые могут играть роль измерений и параметры, которые могут играть роль анализируемых мер;
- продумать содержание аналитических отчетов на основе предполагаемого многомерного представления данных.

Для подготовки к лабораторной работе «Сопровождение и модификация хранилища данных» студенту необходимо:

- изучить разделы 4.1 и 4.2 методического пособия [1];
- изучите возможные изменения источников данных предметной области индивидуального задания;
- составить и согласовать с преподавателем примерный план проверок состояния источников оперативных данных.

Для подготовки к лабораторной работе «Построение функциональной модели процесса разработки хранилища данных» студенту необходимо:

- изучить главы 3 и 4 учебного пособия [1] и раздел 2 методических указаний [3], посвященный построению функциональных моделей;
- выбрать и согласовать с преподавателем какие данные, из собранных вами, согласно индивидуальному заданию, могут служить основой для моделирования отдельных этапов и всего процесса разработки хранилища данных в вашей предметной области.

3.5 Выполнение индивидуального задания

Цель индивидуального задания

Получение навыков комплексного использования изучаемых информационных технологий проектирования и использования хранилищ данных для в конкретной предметной области, использующей современные цифровые технологии организации данных для их анализа в интересах поддержки принятия решений.

Порядок выполнения и содержание работы

1. Студенту необходимо выбрать предметную область и выделить задачи анализа данных, встающие в процессе развития составляющих ее объектов, управление которыми может быть эффективно на основе концепции хранилищ данных.

2. Выделить сферы деятельности в предметной области, являющиеся точками формирования оперативных данных.

3. Собрать данные о предполагаемых объектах управления и описания данных, необходимых для обеспечения их оперативного управления.

4. Сформировать на основе собранных данных атрибутивное представление объектов.

5. Сформулировать задачи анализа данных рассматриваемой проблемной области, для решения которых может быть использована технология концептуального многомерного анализа (OLAP).

6. Сформулировать задачи, для решения которых может быть использована технология хранилищ данных.

7. В ходе выполнения лабораторных работ построить функциональные модели основных этапов проектирования и создания хранилища данных для системы поддержки принятия решений в ходе управления процессами и явлениями выбранной предметной области.

8. Сформулировать задачи анализа состояния объектов предметной области и предложить пути их решения на основе концепции хранилищ данных.

9. Сформулировать задачи анализа состояния объектов предметной области и предложить пути их решения на базе использования системы поддержки принятия решений на основе хранилища данных.

10. Результат выполнения работы может быть представлен либо в печатном, либо в электронном виде. Допускается представление результата в виде презентации.

11. В качестве варианта предметной области может быть выбрана предметная область для лабораторной работы.

Рекомендуемые источники

Для подготовки к выполнению индивидуального задания необходимо ознакомиться как с материалами по применению хранилищ данных для решения задач анализа данных в процессе управления реальными социально-производственными процессами. Одним из источников может служить информационный массив сети интернет.

К начальному набору интернет-источников можно отнести следующие:

1. Сайт компании "Интерфейс", одного из ведущих российских поставщиков инструментальных средств и решений для создания корпоративных информационных систем, разработки приложений, управления проекта-ми, реинжиниринга деятельности предприятий, OLAP – <http://www.interface.ru>.
2. Сайт издательства «Открытые системы» <https://www.osp.ru> , содержащий массу актуальных и интересных статей о проблемах разработки и использования хранилищ данных в современных условиях цифровой экономики, как например статья «Семь этапов успешного проекта построения хранилища данных» <https://www.osp.ru/winitpro/2012/06/13033234>

Примеры тематик предметной области

Поддержка принятия решений в процессе деятельности региональной компании по торговле сельхозпродукцией.

Информационное сопровождение управления сетью цветочных магазинов.

Поддержка принятия решений в процессе деятельности компании по оказанию туристических услуг.

Информационное сопровождение управления деятельностью торговой сети.

Управление сетью учреждений сферы здравоохранения.

Муниципальное управление недвижимым имуществом территории.

Поддержка принятия решений в процессе управления сетью торговых представительств крупной автомобильной корпорации.

Управление сетью заправочных станций.

3.6 Подготовка к зачету

Для подготовки к зачету рекомендуется повторить соответствующие тематике разделы учебно-методического пособия [1].

Вопросы к зачету представлены в рабочей программе изучаемой дисциплине, размещенной на образовательном портале ТУСУРа: <https://edu.tusur.ru/>.

3.7 Подготовка к экзамену

Для подготовки к экзамену рекомендуется повторить соответствующие тематике разделы учебно-методического пособия [1].

Вопросы к экзамену представлены в рабочей программе изучаемой дисциплине, размещенной на образовательном портале ТУСУРа: <https://edu.tusur.ru/>.

4 Рекомендуемая литература

1. Жуковский, О. И. Хранилища данных: Учебное пособие [Электронный ресурс] / О. И. Жуковский. — Томск: ТУСУР, 2015. — 165 с. — Режим доступа: <https://edu.tusur.ru/publications/5366>

2. Сенченко, П. В. Базы данных: Методические указания к лабораторным работам и организации самостоятельной работы [Электронный ресурс] / П. В. Сенченко. — Томск: ТУСУР, 2018. — 75 с. — Режим доступа: <https://edu.tusur.ru/publications/8059>

3. Силич, М. П. Моделирование и анализ бизнес-процессов: Методические указания к лабораторным работам [Электронный ресурс] / М. П. Силич. — Томск: ТУСУР, 2018. — 96 с. — Режим доступа: <https://edu.tusur.ru/publications/7754>