ПЗ-4

1. Какие этапы прохождения данных от первичных источников к пользователям Вы знаете?

* Этап извлечения, преобразования и загрузки данных
* Этап накопления, обеспечения готовности данных к использованию
* Этап применения данных, содержащихся в хранилище, и извлекаемых напрямую из первичных источников.

1. В чем состоит идея гибкой архитектуры данных?

- любому пользователю из числа доверенных лиц должна быть обеспечена возможность доступа к любому разрешенному для использования участку данных, которыми располагает предприятие (организация).

1. Назовите свойства открытых систем.

- расширяемость, масштабируемость, переносимость, интероперабельность, способность к интеграции.

1. Каково содержание свойств расширяемости и масштабируемости?

- расширяемость предполагает возможность включения новых или изменения некоторых прикладных функций информационных систем из числа уже реализованных, не изменяя при этом остальные функциональные подсистемы ИС.

- масштабируемость предусматривает применительно к прикладным программам и базам данных, реализуемым на разных прикладных платформах, возможность изменения их количественных характеристик (размерности решаемых задач, числа обслуживаемых пользователей и т.д.) путем настройки параметров, а не путем перепроектирования и программирования заново.

1. Каково содержание свойств переносимости и интероперабельности?

- Переносимость — это возможность перемещения ИС на другие аппаратно-программные платформы в случае их модернизации или замены с наименьшими затратами, сохраняя инвестиции, вложенные в разработку приложений, формирование массивов данных и обучение пользователей.

- Интероперабельность обеспечивает взаимодействие ИС с другими системами при обращении к информационным ресурсам (базам данных, базам знаний) этих систем или решение определённых задач с использованием их вычислительных ресурсов, если собственных ресурсов недостаточно.

1. Каково содержание свойств способность к интеграции и высокой готовности?

- Способность к интеграции обеспечивает объединение нескольких ИС различного назначения в интегрированную многофункциональную ИС.

- Высокая готовность - мера способности системы принимать и успешно выполнять запросы и задания за доступный интервал времени.

1. Чем определяется значение стандартов при создании и применении ИАС?

- Применение стандартов облегчает организацию сбора, представление их в виде системы стандартных показателей в функциональных подсистемах ИАС, позволяет легко вписывать данные из первичных источников в структуру хранилища данных.

1. Какие стандарты описания показателей Вы знаете?

* ISO 9735-88 «EDI for administration, commerce and transport (EDIFACT). Syntax rules.
* ГОСТ 6.20.2-91 «Элементы внешнеторговых данных (ISO 7372-86)

1. В чем смысл стандартов электронного обмена данными?

Смысл заключается в обеспечении стандартизированного обмена транзакционной и другой цифровой информацией; обеспечить возможности программного взаимодействия компьютерных систем различных сегментов, организаций.

1. Какие проблемы разрешаются ETL-процессами?

- Информация из разных источников приводится к единому формату или к группам форматов, фильтруется и структурируется. Выделяется обеспечение необходимого качества данных, в том числе достоверности, согласованности, соответствия установленным ограничениям и бизнес-правилам и т.д. Качество загружаемых и содержащихся в хранилище данных достигается системой организационных и программно-технических мероприятий. К ним относятся: "ручная" проверка разного рода ошибок, несоответствий, например, неодинаковых названий полей с одним смыслом, и автоматическая с применением программных средств.

1. Назовите стадии ETL-процессов.

* Извлечение;
* Структуризация;
* Обработка;
* Пересылка и импорт данных.

1. Какие аспекты характеризуют сбор данных?

- Два основных: структурный и смысловой. В структурном аспекте данные должны приводиться к одному или группе форматов в системе сбора и хранения данных. Смысловой аспект состоит в содержательном наполнении знаковых структур данных.

1. Каковы требования к качеству данных, помещаемых в ИХ?

- Рассматривают фатальные искажения, ошибки, к которым относятся: отсутствие данных в источнике, ошибка в подключении к источнику, проявляющаяся в отсутствии доступа к необходимым данным, проявление системного сбоя операционной системы.

1. Какие методы повышения качества данных используются в процессе их продвижения в информационное хранилище?

* по критичности ошибок в данных (могут или не могут быть загружены) — ошибки в именах полей, типах данных;
* по правильности форматов и представлений данных;
* на соответствие данных ограничениям целостности;
* на уникальности внутренних и внешних ключей;
* по полноте данных и связей;

1. Какие проблемы решаются при приведении данных к единому формату информационного хранилища?

- Решение проблем в виде знаковых структур очень важно при создании структуры хранилища и его поддержании, которое состоит в первоначальной загрузке и дальнейшем его заполнении и актуализации в процессе эксплуатации.

1. В чем состоит смысл семантического разрыва?

- Смысл заключается в том, что в ходе ETL-процессов велика вероятность искажений, сбоев в работе по причине несогласованности в семантике. Такие ситуации называют «семантические разрывы».

1. Какие виды семантических разрывов Вы знаете?

* «Вавилонский»
* Кросспотоковый разрыв
* Асинхронный разрыв

1. Какие источники данных для информационного хранилища Вы можете назвать?

- Simon (Саймон); Inmon (Инмон).

1. Назовите основные принципы построения информационных хранилищ —

правила Инмона.

* Предметная ориентированность;
* Интегрированность;
* Неизменчивость;
* Поддержка хронологии.

1. В чем заключаются требования предметной ориентированности?

- Требования предметной ориентированности заключаются в компоновке пулов информации по определенным предметным областям или целям, обеспечивающим подготовку и принятие соответствующих решений в соответствующей системе DSS или использование в каком-либо другом виде приложения, например, для разработки планов развития предприятия, инвестиционных проектов, реинжиниринга и в любых работах, для выполнения которых необходимы исходные данные, накопленные в информационном хранилище.

1. Раскройте понятие интегрированности данных в ИХ?

- Интегрированность предусматривает сбор и доработку (предварительную обработку) информации по определенной предметной области из различных источников и превращение ее в организованный по заданным правилам, подчиненным определенной цели, массив в виде гиперкуба или системы поликубов информации.

1. В чем состоят требования неизменчивости?

- Неизменчивость состоит в том, что в хранилище данные из различных источников накапливаются, практически не подвергаются каким-либо изменениям, только в случае крайней необходимости, в основном, при обнаружении искажений. Они наращиваются по заданному, четко определенному графику. Пользователь имеет только право чтения информации. Манипуляции с содержимым хранилища доверяются только администраторам.

1. В чем состоит поддержка хронологии в информационном хранилище?

- Поддержка хронологии заключается в обязательности привязки данных ко времени. Информация, содержащаяся в хранилище, рассматривается в историческом аспекте.

1. Какова структура информационного хранилища?

- Данные в хранилище из источников накапливаются в течение определенного периода времени в зоне накопления. В течение этого времени производится работа по обеспечению необходимого качества данных по описанным выше правилам. В процессе перекачки из источников данные трансформируются в единый формат, проверяются их смысловая согласованность, наличие ошибок, проводятся мероприятия по повышению качества данных. При достижении необходимого уровня качества и определенного графиком работы момента времени данные переносятся в зону хранения.

- В зоне хранения они могут быть представлены в виде реляционной или многомерной модели (объектном виде). В случае использования реляционной модели необходимо иметь в хранилище зону объектного представления данных для достижения уровня характеристик системы, соответствующего требованиям, предъявляемым к OLAP-системам.

1. Раскройте содержание концепции централизованного хранилища данных.

- Такой подход означает, что при нескольких источниках информации — операционных базах данных создается единое централизованное хранилище. В первичных источниках данные хранятся в «сыром» — недоработанном виде, то есть в структуре информационного пространства данного источника информации или операционной БД. Вся поступающая в ИХ информация должна быть преобразована в принятую в данном ИХ структуру. Передача данных из операционных БД в ИХ, которая сопровождается доработкой, может быть организована по заданному временному графику и правилам доработки с соблюдением принципов Инмона. Допускаются неожиданные запросы «на лету», что предъявляет более строгие требования к инструментальным средствам ИХ.

1. В чем состоит концепция распределенного хранилища данных?

- Система локальных хранилищ действует в качестве распределенного хранилища. Не исключается и наличие центрального хранилища, но в такой структуре требования к его размерности значительно облегчаются. Эта концепция предусматривает трансляцию каждого запроса к каждому источнику (базе данных), обработку, увязывание, согласование, компоновку извлеченных данных «на лету» и предоставление их пользователю.

1. Что такое автономные витрины данных?

- Автономные витрины данных — подмножество (срез) хранилища данных, представляющее собой массив тематической, узконаправленной информации, ориентированной, например, на пользователей одной рабочей группы или департамента.

1. В чем смысл концепции единого интегрированного хранилища и многих витрин данных?

- Смысл заключается крупном информационном хранилище агрегированной и подработанной информации, которое может удовлетворить потенциальные запросы по отдельным направлениям деятельности.

II

Тесты

1. Этапы прохождения данных из источников — это:
2. Этап извлечения, преобразования и загрузки данных
3. Этап накопления, обеспечения готовности данных к использованию
4. Этап применения данных, содержащихся в хранилище, и извлекаемых напрямую из первичных источников.
5. Идея гибкой архитектуры данных означает, что:

Любому пользователю из числа доверенных лиц должна быть обеспечена возможность доступа к любому разрешенному для использования участку данных, которыми располагает предприятие (организация).

1. Открытая система согласно определению POSIX 1003.0 принятому Комитетом IEEE это:

Обладающая специальными свойствами система, позволяющая пользователям переходить от системы к системе с переносом данных и программных приложений;

1. Выделите из приведенных свойств систем необязательные для открытых систем свойства:
   1. минимальное время отклика;
   2. многомерность;
   3. поддержка хронологии;
2. Стандарты при создании информационно-аналитических систем применяются для:

Их применение облегчает организацию сбора, представление их в виде системы стандартных показателей в функциональных подсистемах ИАС позволяет легко вписывать данные из первичных источников в структуру хранилища данных.

1. Стандарты электронного обмена данными в англоязычной терминологии Electronic Data Interchange (EDI) предназначены для:

Замены обмена информацией и документами, осуществляемый на бумажных носителях, электронным документооборотом между компьютерными сетями.

1. ETL (Extraction, Transformation, Loading) -процессы сбора, преобразования и загрузки данных обеспечивают:

Подготовку результатов анализа;

1. Сбор данных характеризуют следующие аспекты:
   1. структурный;
   2. смысловой.
2. В процессе продвижения данных в информационное хранилище используются следующие критерии оценки качества данных по структурному представлению:

На кроссязыковый разрыв;

1. Нарушения смыслового содержания данных (семантические разрывы) бывают следующих видов:

Разрыв кодограмм;

1. Основные принципы построения информационных хранилищ — правила Инмона следующие:

Многомерность;

1. Структура информационного хранилища содержит следующие функциональные модули:

Зона хранения данных в многомерном (объектном) виде.

1. Концепция информационного хранилища может быть реализована в нескольких вариантах:
2. корпоративное хранилище данных;
3. База метаданных или репозиторий информационного хранилища представляет собой…

База метаданных — это субканальная информация об используемых данных. Структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления ими.

Репозиторий информационного хранилища — это цифровой репозиторий, который хранит и защищает информацию в компьютерных системах. Оно может быть сетевым хранилищем, распределенным облачным хранилищем, физическим жестким диском или виртуальным хранилищем.