|  |
| --- |
| **МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ** |
|  |
| **НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»** |
| **ФАКУЛЬТЕТ ОЧНО-ЗАОЧНОГО (ВЕЧЕРНЕГО) ОБУЧЕНИЯ НИЯУ МИФИ**  **КАФЕДРА №71**  **«ЭКОНОМИКА И МЕНЕДЖМЕНТ В ПРОМЫШЛЕННОСТИ»** |

**Направление подготовки:**

**38.04.05.Бизнес- информатика**

**Группа М22-В03**

|  |
| --- |
| **Магистерская программа**  **БИЗНЕС-ИНФОРМАТИКА В ЦИФРОВОЙ ЭКОНОМИКЕ** |

**Научно-практический семинар**

**Реферат на тему:**

**Теоретические аспекты построения хранилищ данных**

1. Магистр группы М22-В03 \_\_\_\_\_Хандуев\_Цыден\_Баясхалович\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

(индекс) (фамилия, имя, отчество) (подпись)

2.Руководитель\_\_\_\_\_\_\_\_Фазульянов Дмитрий Владимирович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(фамилия, имя, отчество, должность, кафедра, степень, звание)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.\_\_\_\_\_\_\_\_\_\_\_\_

МОСКВА 2023 г.

Оглавление

[Введение 3](#_Toc136196609)

[Понятие и роль хранилища данных 4](#_Toc136196610)

[Основные свойства Хранилища данных 6](#_Toc136196611)

[Данные хранилища 10](#_Toc136196612)

[Методика (методология) построения хранилищ данных 14](#_Toc136196613)

[Заключение 17](#_Toc136196614)

[Список использованной литературы 18](#_Toc136196615)

## **ВВЕДЕНИЕ**

Хранилища данных являются неотъемлемой частью современных информационных систем и играют ключевую роль в обеспечении эффективного управления и анализа больших объемов данных. В современном информационном обществе объемы данных растут с каждым днем, и организации сталкиваются с необходимостью эффективного хранения, управления и использования этих данных для принятия взвешенных решений.

Актуальность проблемы хранения данных состоит в том, что традиционные методы хранения и управления данными, такие как реляционные базы данных, сталкиваются с ограничениями в обработке и анализе больших объемов данных, неструктурированных данных и данных в реальном времени. В связи с этим возникает потребность в разработке и использовании специализированных хранилищ данных, способных эффективно обрабатывать и анализировать разнообразные типы данных.

Целью исследования является исследование понятия хранилища данных и основных принципов построения хранилищ данных.

Основные задачи исследования:

* Изучение понятия хранилища данных. Это включает анализ основных характеристик и функций хранилищ данных, а также их взаимосвязи с другими компонентами информационной инфраструктуры.
* Анализ основных свойств хранилища данных. Это позволит понять, какие требования и ограничения должны быть учтены при проектировании и использовании хранилищ данных.
* Изучение данных хранилища. Это включает изучение структурированных, неструктурированных и полуструктурированных данных, а также анализ их особенностей и требований к их обработке и хранению.
* Разработка методики построения хранилищ данных. Эта методика будет включать шаги и рекомендации по определению требований к хранилищу данных, анализу модели данных, выбору технологий хранения данных и разработке архитектуры хранилища.
* Оценка эффективности хранилищ данных. Будут исследоваться показатели производительности, надежности, доступности данных и другие факторы, влияющие на качество работы с хранилищем данных.

## **ПОНЯТИЕ И РОЛЬ ХРАНИЛИЩА ДАННЫХ**

Хранилище данных (Datawarehouse) представляет собой централизованное и интегрированное хранилище, предназначенное для хранения и управления большим объемом данных, собранных из различных источников. Оно служит основой для аналитической обработки данных и поддержки принятия управленческих решений в организации.

История возникновения хранилищ данных начинается в конце 1980-х годов. В то время компании столкнулись с проблемой доступа к большим объемам данных, распределенным по различным оперативным системам и базам данных. Традиционные методы хранения и обработки данных не могли обеспечить эффективный анализ и получение ценной информации для принятия стратегических решений.

В 1990-х годах появилась концепция хранилищ данных, предложенная Биллом Инмоном (Bill Inmon) и Ральфом Кимбаллом (Ralph Kimball). Инмон предложил строить хранилище данных как централизованную историческую базу данных, в которой данные интегрируются и преобразуются согласно общему бизнес-пониманию. Кимбалл же предложил подход, основанный на построении звездной схемы (star schema) или снежинки (snowflake schema), где данные организованы вокруг центральной таблицы фактов и связанных с ней измерений.

С появлением хранилищ данных стало возможным собирать, хранить и анализировать большие объемы данных из различных источников, что позволило организациям получать ценные инсайты, принимать обоснованные решения и лучше понимать свой бизнес.

С течением времени технологии хранилищ данных продолжали развиваться, включая появление инструментов для извлечения, преобразования и загрузки данных (ETL), а также интеграцию с бизнес-интеллектом (Business Intelligence) и системами аналитики.

Хранилища данных играют важную роль в современных организациях, предоставляя централизованное и интегрированное хранилище для хранения и управления большим объемом данных. Вот некоторые из ключевых ролей, которые выполняют хранилища данных:

1. Централизация данных: Хранилище данных обеспечивает централизованное место для хранения данных из различных источников. Оно интегрирует данные из разных систем и баз данных, позволяя организации иметь единую и консистентную версию данных. Это способствует повышению надежности, доступности и целостности данных.
2. Историческое хранение данных: Хранилище данных предоставляет возможность сохранять исторические данные, позволяя анализировать прошлые события и тренды. Это позволяет организациям принимать информированные решения на основе анализа данных прошлого периода.
3. Аналитика и отчетность: Хранилище данных служит основой для аналитической обработки данных и создания отчетов. Оно предоставляет возможность проводить сложный анализ данных, выполнять агрегацию, фильтрацию, свертку и другие операции для получения ценных инсайтов. Благодаря хранилищу данных организации могут легко создавать отчеты, дашборды и аналитические инструменты для поддержки принятия решений.
4. Поддержка принятия решений: Хранилище данных обеспечивает доступ к актуальным и своевременным данным, что помогает руководителям и управленцам принимать обоснованные решения. Они могут анализировать текущую ситуацию, оценивать производственные показатели, тренды рынка, поведение клиентов и другие факторы, чтобы разрабатывать эффективные стратегии и тактики.
5. Поддержка управления данными: Хранилище данных предоставляет средства для управления данными, включая механизмы контроля качества данных, защиту и безопасность данных, управление доступом и правами пользователей, а также механизмы резервного копирования и восстановления данных. Это позволяет организациям обеспечивать надежность и конфиденциальность данных, а также соблюдать требования законодательства и регулятивных органов.

## **ОСНОВНЫЕ СВОЙСТВА ХРАНИЛИЩА ДАННЫХ**

*Ориентированность на предметную область:*

Хранилище данных отличается от прикладных приложений своей ориентацией на предметную область. В то время как приложения фокусируются на функциональности и процессах, хранилище данных организовано вокруг фактов и предметов. Например, вместо операций, таких как открытие сделки или кредитование, данные в хранилище описывают сделки, суммы кредитов, покупателей, поставщиков, продукты и т.д.

*Интегрированность:*

Одним из основных аспектов хранилища данных является его интегрированность. Это проявляется в согласованности имен, единиц измерения переменных, структур данных и физических атрибутов данных. В отличие от прикладных приложений, где могут быть использованы разные средства разработки и способы построения приложений, хранилище данных обеспечивает согласованность и единообразие данных.

Интеграция данных по единицам измерения атрибутов:

При разработке приложений разные разработчики могут использовать разные способы задания размеров продукции, например, в сантиметрах, дюймах или ядрах. Однако, в хранилище данных, все данные должны быть приведены к одним и тем же единицам измерения, которые являются стандартом в хранилище. Это обеспечивает согласованность и унификацию данных внутри хранилища.

Существует несколько причин, которые могут привести к рассогласованию приложений. Первая причина связана с использованием различных средств разработки. Каждое средство разработки имеет свои собственные правила, часть из которых является уникальной. Каждый разработчик предпочитает определенные средства разработки, и если два разработчика используют разные инструменты, они могут применять индивидуальные особенности каждого инструмента, что может привести к несогласованности между создаваемыми системами.

Вторая причина рассогласования приложений связана с разнообразием способов построения приложений. Способ разработки конкретного приложения зависит от стиля разработчика, времени разработки и других факторов, характеризующих условия разработки. Это отражается в различных подходах к определению ключевых структур, методах кодирования, обозначении данных, а также в физических характеристиках данных. Поэтому, если два разработчика используют разные способы построения приложений, существует высокая вероятность недостаточной согласованности между системами.

Важным аспектом интеграции данных является согласование единиц измерения атрибутов. Разработчики приложений могут использовать разные способы задания размеров продукции, таких как сантиметры, дюймы, ядра и т.д. Однако, при передаче данных в хранилище, все значения должны быть приведены к одним стандартным единицам измерения, которые приняты в хранилище данных.

*Зависимость от времени*

В хранилище данных все значения данных в определенный момент времени согласованы и не противоречат друг другу. Для оперативных систем эта особенность означает, что данные должны иметь согласованные значения только в момент доступа к ним в оперативной среде.

Зависимость от времени в хранилище данных проявляется следующим образом. Данные в хранилище представлены за определенный временной период, обычно от года до 10 лет. В оперативной среде данные представлены в более коротком временном интервале, обычно от текущего значения до нескольких дней. Для высокопроизводительных приложений, которые обрабатывают много транзакций, важно работать с минимальным объемом данных. Поэтому оперативные приложения ориентированы на более короткий временной период.

Еще одно проявление зависимости хранилища данных от времени заключается в его структуре. Каждая структура хранилища, явно или неявно, включает элемент времени.

Третье проявление зависимости хранилища данных от времени состоит в том, что данные, которые были записаны в хранилище корректно, не могут быть обновлены. Хранилище данных представляет собой последовательность моментальных снимков данных. Если моментальный снимок данных был сделан некорректно, он может быть изменен. Однако, если был получен корректный моментальный снимок, то он не может быть изменен в будущем. С другой стороны, оперативные данные, которые были корректны в момент доступа к ним, могут быть обновлены по мере необходимости.

*Постоянство*

Четвертой важной характеристикой хранилища данных является постоянство. В оперативной среде регулярно происходят операции обновления, добавления, удаления и изменения данных. Однако в хранилище данных не происходит обновление данных. Исходные данные, после того как они были согласованы, проверены и внесены в хранилище данных, остаются неизменными и используются только для чтения.

Существуют важные последствия различий между обработкой данных в оперативной среде и обработкой данных в хранилище данных. На уровне проектирования хранилища данных не требуется поддерживать механизмы, обеспечивающие корректность обновлений, так как обновления не производятся в хранилище данных. Это означает, что физическое проектирование может оптимизировать доступ к данным без ограничений, связанных с обновлениями. Еще одно последствие простоты работы с данными в хранилище данных связано с технологией обработки данных. Технология обработки данных в оперативной среде более сложная, она поддерживает функции резервного копирования и восстановления, обеспечивает целостность данных и включает механизмы разрешения конфликтов и блокировок. Для обработки данных в хранилище данных эти функции не являются столь критическими.

Характеристики хранилища данных, такие как ориентированность на предметную область при проектировании, интеграция данных, зависимость от времени и простота управления данными, создают среду, которая отличается от классической транзакционной среды.

Оперативные среды являются источником практически всех данных в среде хранилища данных. Может показаться, что данные в обеих средах избыточны. Однако на практике избыточность данных в минимальна по следующим причинам:

* При передаче данных из оперативной среды в хранилище данных происходит их фильтрация. Многие данные никогда не выгружаются из оперативной среды, и в хранилище данных передается только информация, необходимая для принятия решений в системе поддержки.
* Временные горизонты в средах также существенно отличаются. Данные в оперативной среде всегда актуальны, в то время как данные в хранилище имеют историческую природу и хронологию. Пересечение данных между оперативной средой и средой хранилища данных минимально с точки зрения временного горизонта.
* Хранилище данных содержит агрегированные итоговые данные, которые никогда не включаются в оперативную среду.
* Передача данных из оперативной среды в хранилище данных сопровождается фундаментальными преобразованиями, и большинство данных при переносе в хранилище изменяется.

## **ДАННЫЕ ХРАНИЛИЩА**

В общем случае модель данных современных Систем Поддержки Принятия Решений строится на основе пяти классов данных:

• источники данных,

• хранилища данных (в узком смысле),

• оперативный склад данных,

• витрины данных,

• метаданные.

*Источники данных*

В качестве источников данных для хранилища используются оперативные транзакционные системы, которые записывают повседневную учетную деятельность компании. Решение о включении конкретной транзакционной системы в качестве источника зависит от бизнес-требований системы поддержки принятия решений (СППР). В соответствии с этими требованиями также могут рассматриваться внешние системы, включая интернет. Детальные данные из источников могут поступать в хранилище данных напрямую или предварительно агрегироваться до нужного уровня обобщения.

*Хранилище данных (в узком смысле)*

В узком смысле хранилище данных представляет собой специализированную базу данных или набор баз данных, которые извлекаются из источников и организованы по сегментам, отражающим конкретную предметную область бизнеса, такие как производство, правила и детальные слабо агрегированные данные.

*Оперативный склад данных (Operational Data Store - ODS)*   
В литературе существуют различные определения оперативного склада данных. В частности, оперативный склад данных может быть рассмотрен как технологический компонент системы поддержки принятия решений (СППР), который служит промежуточным буфером между транзакционными источниками данных и хранилищем. Как упоминалось ранее, данные должны быть преобразованы в единые форматы, очищены, объединены и синхронизированы, прежде чем они попадут в хранилище. Например, данные, необходимые для поддержки процесса принятия решений, могут существовать в транзакционной системе в более короткие сроки (несколько часов, дней), чем период пополнения данных в хранилище (несколько дней, недель). Также возможна ситуация, когда семантически однородные данные поступают из транзакционных систем в разное время. В таких случаях оперативный склад данных служит аккумулятором данных, которые поступают от источников до их загрузки в хранилище. В отличие от хранилища данных, информация в оперативном складе данных может изменяться со временем в соответствии с изменениями, происходящими в источниках данных.

Оперативный склад данных создается в качестве промежуточного буфера между оперативными системами и хранилищем данных. Подобно хранилищу данных, оперативный склад данных также является предметно ориентированным и содержит детальные данные. Отличия оперативного склада данных от хранилища состоят в следующем:

* Изменяемое содержимое.
* Содержит только детальные данные.
* Содержит текущие значения данных.

Детальные данные представляют собой данные из оперативных и внешних систем, которые не подвергаются операциям обобщения или суммирования, то есть данные, сохраняющие свою семантику. Из оперативных систем и внешних источников данные поступают в оперативный склад, проходя процесс трансформации.

Данные оперативного склада регулярно обновляются. Каждый раз, когда данные изменяются в оперативных системах и внешних источниках, соответствующие им данные из оперативного склада также должны быть изменены. Частота обновления оперативного склада зависит как от частоты обновления источников, так и от регламента загрузки данных в склад.

*Витрины данных (Data Mart)*

Функционально ориентированные витрины данных представляют собой структуры данных, предназначенные для решения аналитических задач в конкретных функциональных областях или подразделениях компании, таких как управление прибыльностью, анализ рынков, ресурсов и других. Иногда такие структуры хранения данных также называют киосками данных. Витрины данных можно рассматривать как небольшие хранилища, созданные для информационной поддержки аналитических задач в конкретных управленческих подразделениях компании.

Обычно витрина содержит гораздо меньше данных, охватывает только несколько предметных областей и имеет более ограниченную историю. Витрины данных могут быть представлены в виде логически или физически отделенных подмножеств хранилищ данных. Они обычно строятся для удовлетворения потребностей определенной группы пользователей.

Источником данных для витрин служат данные из хранилищ данных, которые, как правило, агрегируются и консолидируются на разных уровнях иерархии. Детальные данные могут быть включены в витрину или присутствовать в виде ссылок на данные из хранилища.

Различные витрины данных содержат различные комбинации и выборки тех же детализированных данных из хранилища. Важно отметить, что данные в витрине поступают из центрального хранилища данных, которое является единственным "источником истины".

*Метаданные*   
Метаданные представляют собой информацию о данных и играют важную роль в разработке Систем Поддержки Принятия Решений (СППР). Они являются сложными и недостаточно исследованными объектами. Обычно в системе метаданные включают как минимум три аспекта.

С точки зрения пользователей:

* + Метаданные для бизнес-аналитиков.
  + Метаданные для администраторов.
  + Метаданные для разработчиков.

С точки зрения предметных областей:

* + Структуры данных хранилища.
  + Модели бизнес-процессов.
  + Описания пользователей.
  + Технологические аспекты и другие.

С точки зрения функциональности системы:

* + Метаданные о процессах трансформации данных.
  + Метаданные для администрирования системы.
  + Метаданные о приложениях и представлении данных для пользователей.

Присутствие этих трех аспектов метаданных подразумевает, что прикладные пользователи и разработчики системы будут иметь различное понимание технологических аспектов трансформации данных из источников. Прикладные пользователи сконцентрированы на семантике, составе и периодичности обновления хранилища данными из источников, в то время как разработчики фокусируются на ER-диаграммах, правилах трансформации и интерфейсе доступа к данным источников.

В настоящее время отсутствует единая промышленная технология для проектирования, создания и управления метаданными. Поэтому вопросы, связанные с управлением метаданными, рассматриваются индивидуально для каждого конкретного проекта построения СППР.

## **МЕТОДИКА (МЕТОДОЛОГИЯ) ПОСТРОЕНИЯ ХРАНИЛИЩ ДАННЫХ**

Существуют различные подходы к стратегии построения корпоративного хранилища данных (ХД):

* построение сверху вниз,
* снизу вверх,
* динамическая интеграция данных и др.

Считается, что наиболее эффективным подходом является подход, при котором в процессе разработки и внедрения хранилища данных осуществляется его пошаговое наращивание на основе единой системы классификаторов и общей среды передачи и хранения данных – спиральная модель процесса разработки.

|  |  |
| --- | --- |
| постановка  задачи  проекти-  рование  реализация  внедр-  ение  постановка задачи  проектирование  реализация  внедрение  постановка  задачи  проекти-  рование  реализация  внедр-  ение |  |
| Рис. 1 Спиральная модель разработки | Рис. 2 Стратегия построения СППР |

В ходе каждого этапа развертывания происходит реализация одной или нескольких витрин данных в соответствии с технологическим циклом, который включает следующие стадии создания: постановку задачи, проектирование, реализацию и внедрение. Постепенное наращивание системы по этапам позволяет в кратчайшие сроки внедрить законченную систему с ограниченной функциональностью в промышленную эксплуатацию. Масштабы каждого проектного цикла ограничены, что существенно снижает потери при возможных ошибках в проекте по сравнению с полномасштабным проектированием и созданием системы в целом. Поскольку в каждом цикле применяются одинаковые методологические и технологические подходы, а также используются одни и те же средства разработки, время реализации каждой новой витрины будет сокращаться благодаря повышению опыта проектной группы и постепенной отладке взаимодействия между заказчиком и разработчиком системы.

*Постановка задачи*

На стадии постановки задачи проводится системно-аналитическое обследование, которое начинается с утверждения плана и программы обследования заказчиком. В процессе обследования осуществляются следующие работы: проведение интервью с ключевыми участниками проекта со стороны заказчика и лицами, принимающими управленческие решения; уточнение организационной структуры и организационных рамок проекта; выявление особенностей и недостатков существующих информационных решений; формализация схемы бизнеса компании с учетом функциональных рамок; сбор существующих отчетных материалов и официальных документов, относящихся к реализации проекта. Результатом обследования являются уточненные стратегические и оперативные задачи управления компанией, цели и задачи создания системы.

*Техническое задание*

Техническое задание (ТЗ) является одним из основных документов проекта, который определяет требования к созданию системы проектирования и порядок ее разработки. Обычно, если время разработки системы превышает двенадцать месяцев, то целесообразно вводить приоритетность и, соответственно, начинать разработку системы первого этапа на основе концепции, которая может быть реализована в течение 3 месяцев. В противном случае динамичные условия бизнеса и постоянно совершенствующиеся информационные технологии приведут к тому, что полномасштабная система будет уже устаревшей при ее реализации. Если проект достаточно крупный, то помимо основного ТЗ на всю систему могут разрабатываться отдельные ТЗ на ее компоненты.

*Проектирование*

На этапе проектирования, на основе анализа требований к системе, сформулированных в ТЗ, разрабатываются основные архитектурные решения. Архитектура информационной системы рассматривается в четырех аспектах:

• Логическая архитектура. Представляет собой структуру системы с точки зрения набора базовых классов и их взаимосвязей. Здесь определяются автоматизируемые процессы и функции, необходимые для достижения поставленных целей, которые затем разбиваются на задачи, подлежащие реализации на этапе разработки.

• Архитектура процессов. В отношении системы проектирования, определяет информационное обеспечение системы - состав и содержание процессов преобразования и передачи данных.

• Компонентная архитектура. Определяет архитектуру программного обеспечения системы, разбивая ее на подсистемы и компоненты.

• Техническая архитектура. Описывает физические узлы системы и связи между ними.

*Реализация*

На этапе реализации проекта осуществляется разработка и тестирование компонентов информационного и специального программного обеспечения (ПО) системы в соответствии с разработанной на этапе проектирования архитектурой. Основными результатами этой работы являются: • Сама система в виде общего и специального ПО, а также баз данных. • План внедрения системы, который включает все этапы работы по внедрению системы у заказчика, включая упаковку, доставку, установку на технических средствах заказчика, тестирование и доработку. • Набор тестов, которые должны быть выполнены после установки системы у заказчика. • Пользовательскую документацию и обучающие материалы для пользователей системы.

*Внедрение*

Фаза внедрения состоит в выполнении работ, предусмотренных планом внедрения, разработанным на предыдущем этапе. На этапе развертывания происходит установка системы и ее отдельных компонентов у заказчика. Выполняется первоначальная загрузка необходимых данных в хранилище и проводится опытная эксплуатация системы. Кроме того, на этапе развертывания проводится обучение пользователей и сотрудников службы технической поддержки. Завершением этого этапа является переход к производственной эксплуатации хранилища.

## **ЗАКЛЮЧЕНИЕ**

В реферате рассмотрена методология построения хранилища данных, представлены различные подходы и стратегии, а также подробно описана спиральная модель процесса разработки как наиболее эффективный подход.

Основные этапы процесса включают постановку задачи, проектирование, реализацию и внедрение. Каждый этап имеет свои цели и задачи, и их последовательное выполнение позволяет наращивать хранилище данных пошагово, сокращая потери и ускоряя время внедрения.

Стратегия пошагового наращивания хранилища данных позволяет быстро вводить в промышленную эксплуатацию законченные системы с ограниченной функциональностью, снижая риски и повышая опыт проектной группы. Кроме того, описано системно-аналитическое обследование, которое помогает определить требования и цели создания системы.

Техническое задание является ключевым документом проекта, определяющим требования к системе и порядок ее создания. Проектирование осуществляется с учетом логической, архитектурной, компонентной и технической архитектуры.

На стадии реализации разрабатываются и тестируются компоненты системы, а на стадии внедрения осуществляется монтаж, установка системы у заказчика, обучение пользователей и переход к производственной эксплуатации.

С учетом всей представленной информации, можно сделать вывод, что рассмотренная методология построения хранилища данных представляет собой структурированный и системный подход, который позволяет эффективно разрабатывать и внедрять систему поэтапно, сокращая риски и повышая гибкость процесса.

## **СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ**

1. Архипенков С., Голубев Д., Максименко О. ХРАНИЛИЩА ДАННЫХ. От концепции до внедрения - М.: ДИАЛОГ-МИФИ, 2002.
2. Спирли, Эрик. Корпоративные хранилища данных. Планирование, разработка, реализация. Том 1. – М. : Издательский дом «Вильямс», 2001.
3. M.Lea Shaw Data Warehouse Database Design. Student guide - Oracle Corporaton, 2001
4. Richard A.Green Oracle iDS Implement Warehouse Builder. Student guide - Oracle Corporaton, 2001 .
5. Когаловский М.Р. Энциклопедия технологий баз данных. Эволюция и стандарты. Инфраструктура. Терминология. – Москва: "Финансы и статистика", 2002.
6. Шпеник М., Следж О. и др. Руководство администратора баз данных. Microsoft SQL Server 7.0. – Москва-Санкт-петербург-Киев: "Вильямс", 1999.
7. IDC: Data Warehousing Tools: Market Forecast and Analysis, 2000-2004.
8. http://www.nsau.edu.ru/spravki/textbook/modules/stdatmin.html#mining