БУ ВО

«СУРГУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Ханты-Мансийского автономного округа – Югры»

Кафедра автоматизированных систем обработки информации и управления (АСОИУ)

РЕФЕРАТ

по теме: «Распределенные базы данных»

Выполнил студент гр. 606-71м

Бажаев Арман Бейсембаевич

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_(подпись)

Преподаватель:

Горбунов Дмитрий Владимирович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_(подпись)

Сургут, 2019

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc535243941)

[1.ПОНЯТИЕ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ 4](#_Toc535243942)

[2. РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ 6](#_Toc535243943)

[3. ЭТАПЫ ПРОЕКТИРОВАНИЯ РАСПРЕДЕЛЕННОЙ БД 8](#_Toc535243944)

[4. ТИПЫ РАСПРЕДЕЛЕННЫХ СУБД 12](#_Toc535243945)

[ЗАКЛЮЧЕНИЕ 15](#_Toc535243946)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 16](#_Toc535243947)

# ВВЕДЕНИЕ

База данных – совокупность взаимосвязанных данных, организованная по определенным правилам и принципам хранения и манипулирования информацией. База знаний – организованная совокупность знаний, относящейся к некоторой предметной области[1].

По технологии обработки данных БД подразделяются на централизованные и распределенные. Централизованная база данных хранится в памяти одной вычислительной системы. Если эта вычислительная система является компонентом сети ЭВМ, возможен распределенный доступ к такой базе. Система централизованных баз данных с сетевым доступом имеют следующие архитектуры систем: файл-сервер, клиент-сервер.

Распределенная база данных – база данных, включающая фрагменты из нескольких баз данных, которые располагаются на различных узлах сети компьютеров, и, возможно управляются различными СУБД. Одними из достоинств распределенных баз данных является гибкое взаимодействие локальных БД, функциональные возможности централизации узлов, возможность распределения файлов в соответствии с их активностью.

# 1.ПОНЯТИЕ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ

С точки зрения пользователя пользователей и прикладных программ распределенная база данных – обычная локальная база данных.

Основная задача систем управления распределенными базами данных является обеспечение средства интеграции локальных баз данных, располагающихся в некоторых узлах вычислительной сети, с тем, чтобы пользователь, работающий в любом узле сети, имел доступ ко всем этим базам данных как к единой базе данных.

Выделяют следующие типы распределенных БД – однородные и неоднородные. Однородная распределенная БД управляются одной и той же СУБД. В неоднородной локальные базы данных могут относиться к разным моделям данных.

Определяют следующие свойства распределенной БД:

1. локальная автономия: управление данными на каждом из узлов распределенной системы выполняется локально;
2. независимость узлов: все узлы равноправны и независимы, а расположенные на них базы являются равноправными поставщиками данных в общее пространство данных;
3. непрерывные операции: возможность непрерывного доступа к данным;
4. прозрачность расположения: пользователю, обращающий к распределенной БД, не требуется знать о реальном, физическом размещении данных в узлах информационной системы;
5. прозрачная фрагментация: возможность распределенного размещения данных, логически представляющих собой единое целое (горизонтальная фрагментация – хранение строк таблицы на различных узлах, вертикальная фрагментация – распределение столбцов таблицы по нескольким узлам);
6. прозрачное тиражирование: перенос изменений между базами данных техническими средствами, невидимыми пользователю распределенной системы;
7. обработка распределенных запросов: возможность выполнения операций выборки над распределенной базой данных, сформулированных в рамках обычного запроса на языке SQL;
8. обработка распределенных транзакций: обновление распределенной базы данных, не разрушающее целостность и согласованность данных;
9. независимость от оборудования: в качестве узлов распределенной системы могут выступать компьютеры любых моделей и производителей;
10. независимость от операционных систем: многообразие операционных систем, управляющих узлами распределенной системы [2].

Работу с распределенными базами данных обеспечивают распределенные системы управления баз данных. Распределенная система управления баз данных (РСУБД) - комплекс программ, предназначенный для управления распределенной базой данных и позволяющий сделать распределенность информации «прозрачной» для конечного пользователя. Из определения РСУБД следует, что распределенная база данных состоит из нескольких фрагментов, которые могут размещаться на нескольких компьютерах, расположенных в сети и к ней возможен параллельный доступ нескольких пользователей. Назначение обеспечения «прозрачности» состоит в том, чтобы распределенная система внешне вела себя точно так же, как и централизованная. Такое распределение данных позволяет, например, хранить в узле сети те данные, которые наиболее часто используются в этом узле. Такой подход облегчает и ускоряет работу с этими данными и оставляет возможность работать с остальными данными базы данных, хотя для доступа к ним требуется потратить некоторое время на передачу данных по сети.

# 2. РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

Основной целью системы распределенных баз данных является обеспечение управляемого доступа и независимого обращения к данным, распределенным в сети ЭВМ. Под управляемым доступом понимается степень безопасности, необходимая для защиты данных от неавторизованного доступа. Независимость обращения, или разделимость, позволяет пользователям получать доступ к данным через различные вычислительные средства.

Система управления распределенными базами данных обеспечивает средства интеграции локальных баз данных, располагающихся в некоторых узлах компьютерной сети, с тем, чтобы пользователь, работающий в любом узле сети, имел доступ ко всем этим базам данных как к единой базе данных [1]. При этом должны обеспечиваться простота использования системы, возможности автономного функционирования при нарушениях связности сети или при административных потребностях, высокая степень эффективности.

Для клиентских приложений распределенная база данных представляется единым целым. Каждый фрагмент базы данных сохраняется на одном или нескольких компьютерах, которые соединены между собой линиями связи и каждый из них работает под управлением отдельной системой управления базой данных. Пользователь взаимодействует с распределенной базой данных через приложения. Приложения могут быть классифицированы как те, которые требуют доступа к данным на других узлах (локальные приложения), и те, которые требуют подобного доступа (глобальные приложения). В РСУБД должно существовать хотя бы одно глобальное приложение, поэтому любая РСУБД должна иметь следующие особенности:

1. набор логически связанных разделяемых данных;
2. сохраняемые данные разбиты на некоторое количество фрагментов;
3. между фрагментами может быть организована репликация данных;
4. фрагменты и их реплики распределены по различным узлам;
5. узлы связаны между собой сетевыми соединениями;
6. работа с данными на каждом узле управляется локальной СУБД.
7. использование распределенного словаря, содержащего сведения о характере имеющихся данных, их размещении и способе доступа к ним;
8. обеспечение целостности данных при связи между электронными таблицами, находящимися в различных системах;
9. независимость способа и процедуры доступа к файлу от места его хранения;
10. работа с частью базы данных, расположенной в одной системе, не может быть прервана обращением из другой системы;
11. администратор части базы, находящейся в одной системе, работает независимо от администраторов частей базы, расположенных в других системах.
12. принцип дублирования - изменение, вносимое в одну часть базы данных, в течение определенного времени отражается и в других частях базы [2].

СУБД на каждом узле способна поддерживать автономную работу локальных приложений.

# 3. ЭТАПЫ ПРОЕКТИРОВАНИЯ РАСПРЕДЕЛЕННОЙ БД

Проектирование распределенных баз данных является сложным процессом, в реализации которого выделяют четыре основные проблемы:

1. проблему дезагрегации: необходимость рационального, в соответствии с системой расчетов (решаемых задач), распределения учетной информации по уровням обработки и участкам учета с обеспечением их взаимосвязи;
2. создание инфологической структуры информационного фонда распределенной базы данных, ориентированного на решение всего комплекса задач избранной системы расчетов;
3. решение требований рационализации вычислительного процесса на основе распределенной базы данных и распределенного комплекса технических средств;
4. организационно-правовую проблему, состоящую в обеспечении защиты данных н соблюдении юридических норм доступа к базам данных, их заполнения, изменения и уничтожения.

Основные этапы последовательности проектирования распределенной базы данных:

* 1. формулирование и анализ требований;
  2. концептуальное проектирование;
  3. проектирование реализации;
  4. разбиение БД;
  5. размещение БД;
  6. физическое проектирование.

На этапе формулирования и анализе требований устанавливаются цели организации, определяются специфичные требования к базе данных, вытекающие из этих целей или сформулированные непосредственно управляющим персоналом организации. Эти требования документируются в форме доступной как конечному пользователю, так и проектировщику базы данных. Специфичные цели и требования к базе данных необходимо определить на самом высоком уровне организации. Собранные и документированные требования должны включать ограничения, обуславливающие безопасность, надежность, уровень достигнутой технологии, а также политические и бюрократические ограничения.

Этап концептуального проектирования связан с описанием и синтезом разнообразных информационных требований пользователей в первоначальный проект баз данных. Результатом этого этапа является высокоуровневое представление информационных требований, например, такое как диаграмма «сущность-связь». Основу этой диаграммы составляет набор сущностей, который представляет или моделирует определенную совокупность сведений, специфицированную в требованиях. Связи между сущностями отображают функциональные аспекты информации, представленной сущностями.

Существует несколько подходов к построению диаграмм типа «сущность-связь». Общим для всех подходов является набор из четырех основных проектных решений или шагов:

1. Определение сущностей.

2. Определение атрибутов сущностей.

3. Идентификации ключевых атрибута сущностей,

4. Определение связей между сущностями.

После построения первоначальной информационной структуры следует ее уточнение и совершенствование. Главной целью этапа проектирования реализации является создание СУБД-ориентированной схемы с использованием в качестве исходных данных результатов концептуального проектирования и требований обработки (UР-информации).

Вначале анализируется содержание требований обработки данных. Формат локальных информационных структур, подлежащих обработке, соответствует формату первоначальной структуры, полученной на этапе концептуального проектирования. После того как представлены все виды обработки, первоначальная структура может быть объединена со всеми локальными структурами в новую информационную структуру. Затем, используя знания, полученные в процессе пересмотра и объединения информационных структур, учитывая связи обрабатываемых данных и характеристики типов записей, допускаемых СУБД, можно сформировать предварительные типы записей.

Построенная таким образом логическая структура базы данных (схема) может быть оценена количественно с помощью таких характеристик, как число обращении к логическим записям, объем обрабатываемых в каждом приложении данных, общий объем хранимых данных[3].

Этап расчленения базы данных связан с расчленением глобальной базы данных и синтезом различных приложений на основе модели. Существуют три класса выходных данных этапа расчленения: 1) совокупность расчлененных частей базы данных (разделов), 2) размер каждого раздела. 3) модели и частоты использования приложений. На этом этапе проектирования исходная глобальная база данных расчленяется на множество подфайлов, которые содержат в точности все сведения, имевшиеся в глобальной базе данных. Каждый подфайл в расчлененной базе данных выбирается как неделимая единица размещения данных. Далее проводится анализ того, как приложения базы данных используют возможные разделы базы данных. Связь между разделом базы данных и приложениями характеризуется идентификатором типа приложения, идентификатором узла сети, создающего приложение, частотой использования приложения и моделью приложения.

Размещение распределенной базы данных является многовариантной задачей. Количество возможных вариантов реализации расчлененной или смешанной базы данных велико. Для выбора наиболее подходящей стратегии распределения данных, необходимо еще до выбора СУБД провести оценку пользовательских и системных требований.

Физическое проектирование базы данных заключается в расширении ее логической модели такими характеристиками, которые необходимы, во-первых, для определения способов физического хранения и использования базы данных и, во-вторых, для определения объемов памяти, требуемой для всей системы и для оценки эффективности обработки. Подобные характеристики касаются того, как и где хранить данные, как их можно найти и использовать.

# 4. ТИПЫ РАСПРЕДЕЛЕННЫХ СУБД

Распределенные СУБД классифицируются как гомогенные и гетерогенные. Если все узлы распределенной системы используют один и тот же тип СУБД, то такая система называется гомогенной. В гетерогенных системах на сайтах могут функционировать различные типы СУБД, использующие разные модели данных.

Гомогенные системы проще проектировать и сопровождать. Такие РСУБД позволяют наращивать размеры системы, добавляя узлы к существующей распределенной системе. Гетерогенные системы возникают в случаях интеграции во вновь создаваемую распределенную систему независимых узлов со своими собственными системами баз данных. В гетерогенных системах для организации взаимодейтсвия между различными типами СУБД потребуется организовать трансляцию передаваемых сообщений. Пользователи каждого из узлов должны иметь возможность вводить свои запросе на языке той СУБД, которая используется на этом узле. Система должна обеспечить локализацию требуемых данных и выполнение транслции передаваемых сообщений. В общем случае данные могут быть запрошены с другого узла, который имеет:

1. иной тип используемого оборудования;
2. иной тип используемой СУБД;
3. иной тип применяемого оборудования и СУБД [3];

Если используется другой тип оборудования, но та же СУБД, методы выполнения трансляции заключаются в замене кодов и изменении длины слова. Если типы используемых на узлах СУБД различны, трансляция усложняется необходимостью отображения структуры данных одной модели в соответствующие структуры данных другой модели. Например, отношения в реляционной модели данных должны быть преобразованы в записи и наборы, типичные для сетевой модели данных. Кроме того, необходимо транслировать текст запросов с одного используемого языка манипулирования данными в другой.

Дополнительные сложности возникают при попытке выработки единой концептуальной схемы, создаваемой путем интеграции независимых локальных концептуальных схем.

Типичное решение, применяемое в некоторых реляционных системах для обеспечения прозрачности в отношении используемой СУБД, состоит в использовании шлюзов. В состав отдельных частей гетерогенных распределенных систем должны входить шлюзы, предназначенные для преобразования языка и модели данных каждого из используемых типов СУБД в язык и модель данных реляционной системы. Такой подход имеет некоторые недостатки: шлюзы не позволяют организовать систему управления транзакциями даже для отдельных пар систем. То есть шлюз между двумя системами представляет собой не более чем транслятор запросов.

Созданная рабочая группа должна подготовить спецификации, регламентирующие инфраструктуру среды базы данных:

1. унифицированный интерфейс языка SQL, позволяющий создавать клиентские приложения так, чтобы они не были привязаны к конкретному типу используемой СУБД;
2. унифицированный протокол доступа к базе данных, позволяющий СУБД одного типа непосредственно взаимодействовать с СУБД другого типа, без необходимости использования какого-либо шлюза;
3. унифицированный сетевой протокол, позволяющий осуществлять взаимодействие СУБД различного типа.

Наиболее важной задачей этой группы является поиск способа, позволяющего в одной транзакции выполнять обработку данных, содержащихся в нескольких базах, управляемых СУБД различных типов, без необходимости использования различных шлюзов.

Мультибазовая система – распределенная система управления базами данных, в которой управление каждым из узлов осуществляется автономно.

В мультибазовых системах предпринимается попытка интеграции таких распределенных систем баз данных, в которых весь контроль над отдельными локальными системами целиком и полностью осуществляется их операторами. Полная автономия узлов позволяет не вносить какие-либо изменения в локальные СУБД [3].

Мультибазовые системы позволяют конечным пользователям разных узлов получать доступ и совместно использовать данные без необходимости физической интеграции существующих баз данных. Они обеспечивают пользователям возможность управлять базами данных их собственных без какого-либо централизованного контроля. Администратор локальной базы данных может разрешить доступ к определенной части своей базы данных посредством создания схемы экспорта.

# ЗАКЛЮЧЕНИЕ

Распределенные базы данных применяются в масштабных проектах. Подобное распространение стало возможным благодаря надежности, при которой стало доступным обеспечить доступ к необходимым данным практически в любой момент времени; высокой скорости выполнения операций с большими объемами данных; возможности совместного использования удаленных данных.

Системы РБД характеризуются жесткими требованиями по структуре, что в значительной степени ограничивает их распространение, а также сложностью распределенных систем с технической точки зрения.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Распределенные базы данных [Электронный ресурс] [2018] – Режим доступа: <https://ru.wikipedia.org/wiki/Распределенные_базы_данных>
2. Распределенные и параллельные системы баз данных [Электронный ресурс] [2018] Режим доступа: <https://citforum.ru/database/classics/distr_and_paral_sdb/>
3. Бойко В.В., Савинков В.М. Проектирование баз данных информационных систем. - М.: Финансы и статистика, 1989. - 351 с.