**Лекция 2. Основные понятия распределенных баз данных.**

1. **Понятие распределенной системы баз данных.**
2. **Обработка данных по технологии клиент-сервер.**

**3. Моделирование распределенных СУБД.**

**4. Распределенная обработка данных в распределенных СУБД.**

**1. Распределенная система баз данных.**

Распределенная База Данных (РБД) или Distributed DаtaBase (DDB) , **это совокупность распределенных данных представленных в рамках некоторой реляционной модели отражающей данные и связи между ними**

Реляционная модель данных обеспечивает однородность представления данных и связей между ними в виде концептуальной схемы базы данных. Реализация реляционной модели приводит к созданию распределенной базы данных.

Сущность РБД заключается в организации доступа пользователей к большим объемам информации. Это позволяет располагать данные так, что последние, с одной стороны, находятся в пунктах наибольшего их спроса, а, с другой стороны, с помощью транзакций обеспечивается доступ к любым данным, не зависимо от того, где они находятся.

РБД, создаваемая заново, является однородной. Вместе с этим, нередко она образуется как совокупность группы локальных баз данных, уже функционирующих в ряде систем. В этом случае возникает неоднородная РБД. Оба типа баз погружаются в Систему Управления Распределенной Базой Данных (СУРБД).

В общем случае локальные базы данных, составляющие распределенную базу данных, не обязательно должны быть однородными (т.е. вестись одной СУБД) или обрабатываться в среде одной и той же операционной системы и/или на компьютерах одного и того же типа.

Например, одна база данных может быть базой Oracle на компьютере SUN с операционной системой SUN OS(UNIX), вторая база данных может вестись СУБД DB2 на мейнфрейме IBM 3090 с операционной системой MVS, а для ведения третьей базы может использоваться СУБД SQL/DS также на мейнфрейме IBM, но с операционной системой VM. Обязательно только одно условие – ЛБД связанны между собой логически, но физически расположенны на нескольких машинах (узлах), входящих в одну компьютерную сеть.

Важнейшие функциональные характеристики распределенной базы данных таковы:

* автономность узлов распределенной базы данных которая означает, что ведение каждой базы может происходить независимо от других.
* обработка распределенных запросов - таких запросов (SQL-предложение), в ходе выполнения которых происходит доступ к объектам (таблицам или представлениям) разных баз данных.
* выполнение распределенных транзакций. При выполнении распределенных транзакций осуществляется согласованное управление всеми вовлеченными базами данных. При этом часто используется технология двухфазной передачи информации для выполнения распределенных транзакций.

**Структура распределенной базы данных.**

В состав РБД включаются:

* Региональных базы данных. РБД содержат пользовательскую информацию которая формируется в данном регионе и в основном в нем используется. Устанавливается на региональном сервере. Как правило в рамках одной системы структуры РБД совпадают, на в каждом регионе возможно наличие таблиц специфических для данного региона. Например для распределенной кадровой системы фирмы с филиалами набор таблиц- "Сотрудники", "Отделы", "Вакансии". Специфическая для киевского региона "Сотрудники пригорода"
* Главная база данных. ГБД содержит интегрированную информацию пользователя которая необходима для функционирования системы , но которой нет в региональных базах данных. Алгоритмы интегрирования (централизации информации специфичны для предметной области). Например "Резерв компании"
* Метабаза данных. (База метаданных системы). База метаданных содержит:
* Набор таблиц отражающих структуру распределенной базы данных ;
* Набор таблиц необходимых для реализации алгоритмов репликации и фрагментации данных

1. **Обработка данных по технологии клиент-сервер.**

Современные компьютерные технологии при проектировании распределенных баз данных рассматриваются в различных вариантах архитектуры «клиент- сервер» (двухзвенная, трехзвенная), которая ориентируется на объектно-ориентированный подход. В этой архитектуре любой объект, использующий ресурсы другого объекта определяется как клиентом, а объект, поставляющий ресурсы называется сервером в соответствии с установленным протоколом обмена

**Архитектура клиент-сервер.**

Для использования РБД необходимо следующее аппаратное и программное обеспечение:

* компьютер-сервер базы данных;
* компьютер(ы)-клиенты;
* коммуникационная сеть;
* сетевое программное обеспечение;
* прикладное программное обеспечение.

Технология клиент-сервер является реализацией распределенной обработки данных. В системе архитектуры клиент-сервер обработка данных разделена между компьютером-клиентом и компьютером-сервером, связь между которыми происходит по сети. Это разделение процессов обработки данных основано на группировании функций. Как правило, компьютер-сервер баз данных выделяется для выполнения операций с базами данных, а компьютер-клиент выполняет прикладные программы. На рисунке 1 показана простая система архитектуры клиент-сервер, в состав которой входят компьютер, действующий как сервер, и другой компьютер, действующий как его клиент. Каждая машина выполняет различные функции и имеет свои собственные ресурсы.

компьютер-сервер

Приложения- обработчики

Приложения- обработчики

Сервер БД База данных

( операции с БД)

сеть

Клиенты IBM-совместимый ПК IBM-совместимый ПК

**Компьютеры-клиенты**

Основная функция компьютера-клиента состоит в выполнении приложения (интерфейса с пользователем и логики представления) и осуществлении связи с сервером, когда этого требует приложение.

Обычно клиенты - это медленные и недорогие машины, такие как персональный компьютер с процессором 386 и 4-мя Мбайтами оперативной памяти, работающий под DOS или OS/2. В зависимости от требований, предъявляемых приложениями, могут понадобиться более мощные компьютеры-клиенты.

**Компьютер-сервер базы данных**

Как следует уже из самого термина, главная функция компьютера-сервера заключается в обслуживании потребностей клиента. Термин "Сервер" используется для обозначения двух различных групп функций: файл-сервер и сервер баз данных. Файл-серверы не предназначены для выполнения операций с базами данных, их основная функция - разделение файлов между несколькими пользователями, т.е. обеспечение одновременного доступа многих пользователей к файлам на компьютере файл-сервере. Примером файл-сервера является операционная система NetWare компании Novell.

Компьютер с сервером баз данных ( например Oracle ) выделяется только для выполнения операций с базами данных.

Основные функции сервера баз данных.:

* Связь с клиентом, анализ и выполнение запроса к базе данных, включая возврат клиенту результата запроса (набора строк из базы данных)
* управление одновременным доступом к базе данных многих пользователей, перенаправление запросов к другим серверам сети
* выполнение реляционных операций, управление словарем-справочником данных и обеспечение защиты.

Одно из важных требований к серверу -- это то, что операционная система, в среде которой размещен сервер баз данных, должна быть многозадачной (и, желательно, но не обязательно, многопользовательской). Многие разновидности систем UNIX, MVS, VM и некоторые другие операционные системы являются и многозадачными, и многопользовательскими.

При выборе компьютера-сервера базы данных следует принять во внимание, мощность процессора, емкость оперативной памяти, объем дисков, скорость работы с дисками, число одновременно работающих пользователей и стоимость.

При этом сетевая поддержка РБД характеризуется следующими особенностями :

* Каждый узел обладает собственными базами данных;
* Узлы работают согласовано, то есть пользователь может получить доступ к данным на любом узле сети, так как будто бы они находятся на его собственном узле.
* Функционирование РБД происходит независимо от типов используемых в системах устройств.
* Обеспечивается пространственная прозрачность, дающая возможность пользователю не знать, где расположены компоненты базы;
* Поддержка полная функциональность, т.е. возможность выполнения всех тех же операций, которые возможны в базе, находящейся в одной системе;
* Гарантия целостности данных, обеспечиваемая функциями слежения за данными, исправления ошибок;

## Принципы взаимодействия между клиентскими и серверными частями

Доступ к базе данных от прикладной программы или пользователя производится путем обращения к клиентской части системы. В качестве основного интерфейса между клиентской и серверной частями выступает язык баз данных SQL.

Это язык по сути дела представляет собой текущий стандарт интерфейса СУБД в открытых системах. Собирательное название SQL-сервер относится ко всем серверам баз данных, основанных на SQL. Серверы баз данных, интерфейс которых основан исключительно на языке SQL, обладают своими преимуществами и своими недостатками. Очевидное преимущество - стандартность интерфейса. В пределе, хотя пока это не совсем так, клиентские части любой SQL-ориентированной СУБД могли бы работать с любым SQL-сервером вне зависимости от того, кто его произвел.

Недостаток тоже довольно очевиден. При таком высоком уровне интерфейса между клиентской и серверной частями системы на стороне клиента работает слишком мало программ СУБД. Это нормально, если на стороне клиента используется маломощная рабочая станция. Но если клиентский компьютер обладает достаточной мощностью, то часто возникает желание возложить на него больше функций управления базами данных, разгрузив сервер, который является узким местом всей системы.

Одним из перспективных направлений СУБД является гибкое конфигурирование системы, при котором распределение функций между клиентской и пользовательской частями СУБД определяется при установке системы.

**3. Моделирование распределенных СУБД.**

Существуют различные модели для построения схем расположенной базы данных. Благодаря схеме пользователь рассматривает РБД как единую БД.

Различные авторы предлагают различные подходы к построению набора схем разрабатываемых при проектировании баз данных. Одной из наиболее распространенных схем является схема Арсеньева-Яковлева (Интеграция расположенных баз данных, ”ЛАНЬ”, 2001 год)

Даталогическое представление

R – схема для СУБД в виде набора SQL скриптов

Информационное проектирование

Инфологическая схема (модель)

Уточненная концептуальная схема

Концептуальное проектирование

Диаграмма прецедентов

Концептуальная схема (модель)

**2.1. Концептуальное моделирование.**

Состоит в выполнении следующих операций:

* определение источников знаний о предмете (закон эвристики, существующие решения, документы);
* определение назначения и основных функций системы ( построение диаграммы прецедентов)
* выбор системы классификации объектов;
* выделение абстракции объектов и процессов т.е. выделение классов (типизация, абстрагирование);
* построение структуры базовых классов;
* построение схемы взаимодействия.

Цель концептуального построения → построение фундаментальной семантики в представленной области.

Нотации описания различны, например параметр в UML языке концептуальная схеме может быть представлена в виде диаграмм классов без баз детализации по атрибутам и методам.

**2.2. Инфологическое моделирование.**

Выполняется на основе концептуальной модели как её уточнение и детализации.

Основные операции:

* формирование базовых атрибутов классов исходящих из типизации.
* формирование дополнительных атрибутов;
* определение совокупного совместного использования б атрибутов, т.е. порождение прототипов таблиц;
* определение связей логической целостности и связей навигации;
* построение информационно- логической модели.

Нотации различны.

В UML языке - полная диаграмма статических классов.

**2.3. Даталогическое моделирование.**

Даталогический анализ и проектирование это окончательное проектирование реляционной схемы данных учетом среды выбранной СУБД.

Выполнены на основе инфологической схемы и заключается в выполнении:

* выделение таблиц (R-отношений);
* определение физических форматов атрибутов на основе типов СУБД;
* определение состава индексированных полей;
* определение связей навигации и логической целостности;
* установка базнес-правил;
* проектирование представлений;
* проведение нормализации реляционных схем (устранение избыточности, многозначности).

Результат даталогического проектирования - реляционная схема базы данных

**4. Распределенная обработка данных -**  Disdtributed Data Processing (DDP)

Выполнение операций с базами данных на одной машине и приложений на другой и есть распределенная обработка данных. Связь между приложением и процессами выполнения операций с базами данных реализуется как аппаратурой, так и путем программного обеспечения. Сущность DDP заключается в том, что пользователь получает возможность работать с сетевыми службами и прикладными процессами, расположенными в нескольких взаимосвязанных абонентских системах. При этом возможны несколько видов работ, которые он может выполнять:

* удаленный запрос, например, команда, позволяющая посылать одиночную заявку на выполнение обработки данных;
* удаленная трансакция, осуществляющая направление группы запросов прикладному процессу;
* распределенная трансакция, дающая возможность использования нескольких серверов и прикладных процессов, выполняемых в группе абонентских систем.