

Содержание:

image not found or type unknown



Введение

Для принятия обоснованных и эффективных решений в производственной деятельности, в управлении экономикой и в политике современный специалист должен уметь с помощью компьютеров и средств связи получать, накапливать, хранить и обрабатывать данные, представляя результат в виде наглядных документов. Поэтому в данной работе рассмотрим работу с базами данных.

В настоящее время одной из важнейших разновидностей информационных систем являются базы данных, в которых накапливается всевозможная информация, и откуда к ней получают доступ десятки миллионов пользователей по всему миру. Проектирование и создание баз данных, накопление в них информации, поддержка баз в актуальном состоянии, а также организация получения из баз любой нужной информации является очень сложной задачей, которая по плечу только высококвалифицированным специалистам.

База данных- организованная совокупность данных, предназначенная для длительного хранения во внешней памяти ЭВМ и постоянного применения. . Группы объектов, сведения о которых сосредотачиваются в базах данных, принято называть сущностями. Базы данных предоставляют возможность выбирать данные из их общего хранилища различным пользователям с различными запросами и критериями поиска.

Любая рассматриваемая сущность (объект, явление) независимо от ее материальности или идеальности имеет некоторые характерные, неотъемлемые черты, свойства, качества, которые принято называть атрибутами этих сущностей. При этом информационная модель сущности представляет собой конкретный набор атрибутов, которые выбраны для ее описания из всего множества атрибутов.

Основными операциями с базами данных являются: создание, первичное заполнение, последующее внесение всевозможных изменений в данные (например,

сведений о поступлении новых товаров или об изменении цен на них и т.д.), сортировка данных, которая используется с целью ускорения поиска, и, наконец, собственно поиск нужных данных, осуществляемый в разных формах. Для выполнения всех этих операций служат инструментальные пакеты программ, которые называются системами управления базами данных (СУБД). Не следует путать базу данных, представляющую собой хранимую информацию, и систему управления базой данных, которая представляет собой программную систему, обеспечивающую все необходимые для хранения и поиска действия.

Существует огромное количество разновидностей баз данных, отличающихся по разным критериям.

Классификация по модели данных:

Сетевая;

Реляционная;

Функциональная;

Объектно-реляционная.

Классификация по содержанию:

)Историческая;

)Научная;

)Географическая.

Классификация по степени распределённости:

)Централизованная;

)Неоднородная;

)Однородная;

)Распределенная.

В данной работе будет рассматриваться распределенная база данных.

Распределенная базы данных

Одной из главных тенденций в компьютерной отрасли в конце 80-х и 90-х годов был переход от централизованных архитектур к распределенным системам, объединенным в вычислительные сети. С появлением мини-компьютеров такие задачи, как учет складских запасов и обработка заказов, были перенесены с больших корпоративных ЭВМ на вычислительные системы среднего уровня, обслуживающие отдельные подразделения, входящие в состав предприятия. Стремительный рост популярности персональных компьютеров в восьмидесятые годы привел к их появлению на рабочих столах у миллионов людей. Следствием широкого распространения персональных компьютеров стало то, что их начали соединять между собой в локальные вычислительные сети (ЛВС), управляемые аппаратным и программным обеспечением таких компаний, как ZCOM и Novell. Позднее ЛВС отдельных рабочих групп стали объединяться в крупные корпоративные сети с помощью маршрутизаторов и другого сетевого оборудования компаний типа Cisco Systems. С началом эпохи Internet, компьютерные сети получили новый импульс развития. Сегодня Internet представляет собой глобальную информационную среду, объединяющую компьютеры по всему миру и позволяющую людям, находящимся на разных концах земного шара, не только оперативно получать интересующую их информацию, но также обмениваться сообщениями и заниматься электронной коммерцией.

С увеличением числа компьютеров в организациях и появлением компьютерных сетей данные перестали размещать в одной вычислительной системе, работающей под управлением одной СУБД. Вместо этого информацию начали распределять по различным системам, каждая из которых имеет собственную СУБД. Зачастую такие вычислительные системы и СУБД поставляются различными изготовителями. Еще большие трудности создает Internet. Даже если в рамках корпорации используется единая структура баз данных и используется одна общая СУБД, то это не означает, что аналогичных соглашений придерживаются клиенты компании. Описанные тенденции привели к тому, что в компьютерной промышленности внимание специалистов по обработке данных сосредоточились на проблемах управления распределенными базами данных.

Под распределенной обычно понимают базу данных, декомпозированную и фрагментированную на несколько узлов вычислительной сети, с возможным управлением различными системами управления базами данных.

image not found or type unknown



Распределенные базы данных должны обладать:

Локальными и глобальными (распределенными) средствами доступа к данным(СУБД).

Единообразной логикой прикладных программ во всех АРМах сети.

Малым временем реакции на запросы пользователей.

Надежностью, исключающей разрушения целостности системы в случае выхода из строя ее отдельных компонентов (узлов).

Открытостью, позволяющей наращивать объем локальных баз данных и добавлять АРМ.

Развитой системой управления резервным копированием и восстановления данных на случай сбоев.

Защищенностью, следящей за соблюдением привилегий доступа к данным.

Высокой эффективностью, за счет выбора оптимальных алгоритмов использования сетевых ресурсов.

Развитым репликационным механизмом, позволяющим размещать обновленные копии данных в сети оптимальным образом.

Принципы построения распределенных баз данных:

Минимизация интенсивности обмена данными.

Оптимальным размещением серверных и клиентских приложений в сети.

Декомпозиция данных на часто и редко используемые сегменты (для правильной настройки репликации- размещение наиболее часто используемых данных на АРМ конечных пользователей).

Периодическое сохранение копий данных и выполнение действий по поддержке целостности распределенной информационной системы.

Критерии построения распределенных баз данных:

Всесторонний анализ информационных потребностей предметной области с выявлением объемов хранимых данных, их сложности, достоверности, взаимосвязанности.

Моделирование предполагаемого сетевого трафика при работе распределенных баз данных с различными моделями репликации данных.

Кластеризация элементов данных и программ их обработки. Цель- добиться максимальной автономности и слабосвязанности кластеров.

Привязка кластеров данных к вероятным пользователям или АРМ.

Поддержка эталонной копии данных и ограничение репликационного механизма.

Разработка и реализация правил приведения локальных и центральной базы данных в непротиворечивое состояние.

Распределенные архитектуры баз данных принято подразделять по типам на:

Системы недублирующего разбиения (при большом объеме часто меняющихся данных).

Системы частичного дублирования (при небольшом объеме часто меняющихся данных).

Системы полного дублирования (при небольшом объеме редко меняющихся данных).

В заключении сформулируем ряд свойств, которым по К.Дейту должна удовлетворять распределенная база данных:

Локальная автономия- это качество означает ,что управление данными на каждом из узлов выполняется локально. База данных, расположенная на одном из узлов, является неотъемлемым компонентом распределенной системы. Будучи фрагментом общего пространства данных она, в тоже время функционирует как полноценная локальная база данных; управление ею выполняется локально и независимо от других узлов системы.

Независимость узлов - в идеальной системе все узлы равноправны и независимы, а расположенные на них базы являются равноправными поставщиками данных в

общее пространство данных. База данных на каждом из узлов самодостаточна- она включает полный собственный словарь данных и полностью защищена от несанкционированного доступа.

Непрерывные операции- это качество можно трактовать как возможность непрерывного доступа к данным в рамках DDB вне зависимости от их расположения и вне зависимости от операций, выполняемых на локальных узлах. Это качество можно выразить лозунгом «данные доступны всегда, а операции над ними выполняются непрерывно».

Прозрачность расположения- это свойство означает полную прозрачность расположения данных. Пользователь, обращающийся к DDB, ничего не должен знать о реальном, физическом размещении данных в узлах информационной системы. Все операции над данными выполняются без учета их местонахождения. Транспортировка запросов к базам данных осуществляется встроенными системными средствами.

Прозрачная фрагментация- это свойство трактуется как возможность распределенного размещения данных, логически представляющих собой единое целое. Существует фрагментация двух видов: горизонтальная и вертикальная. Первая означает хранение строк одной таблицы на различных узлах. Вторая означает распределение столбцов логической таблицы по нескольким узлам.

Прозрачное тиражирование- тиражирование данных это асинхронный процесс переноса изменений объектов исходной базы данных в базы, расположенные на других узлах распределенной системы. В данном контексте прозрачность тиражирования означает возможность переноса изменений между базами данных средствами, невидимыми пользователю распределенной системы. Данное свойство означает что, тиражирование возможно и достигается внутрисистемными средствами.

Обработка распределенных запросов-это свойство DDB трактуется как возможность выполнения операций выборки над распределенной базой данных, сформулированных в рамках обычного запроса на языке SQL. То есть операцию выборки из DDB можно сформулировать с помощью тех же языковых средств, что и операцию над локальной базой данных.

Типы распределенных баз данных

Распределенные базы данных;

Мультибазы данных с глобальной схемой. Схема мультибаз данных- это распределенная система, которая служит внешним интерфейсом для доступа ко множеству локальных СУБД или структурируется, как глобальный уровень над локальными СУБД.

Федеративные базы данных. В отличие от мультибаз не располагают глобальной схемой, к которой обращаются все приложения. Вместо этого поддерживается локальная схема импорта- экспорта данных. На каждом узле поддерживается частичная глобальная схема, описывающая информацию тех удаленных источников, данные с которых необходимы для функционирования.

Хранилище данных и OLAP. Назначение. Основные характеристики

В основе технологии хранилищ данных лежит идея о том, что базы данных, ориентированные на оперативную обработку транзакций, и базы данных, предназначенные для делового анализа, используются совершенно по-разному и служат разным целям. Первые- это средство производства, основа ежедневного функционирования предприятия. На производственном предприятии подобные базы данных поддерживают процессы принятия заказов клиентов, учета сырья, складского учета и оплаты продукции, т.е. выполняют главным образом учетные функции. С такими базами данных, как правило работают клиентские приложения, используемые клерками, производственным персоналом. В противоположность этому базы данных второго типа используются для принятия решений на основе сбора и анализа информации. Их главные пользователи- это менеджеры, служащие планового отдела и отдела маркетинг.

Хранилище данных- предметно-ориентированный , интегрированный, привязанный ко времени и неизменяемый набор данных, предназначенный для поддержке принятия решений.

Хранилище данных содержит непротиворечивые, консолидированные, исторические данные и предоставляет инструментальные средства для их анализа с целью поддержки принятия стратегических решений. Информационные ресурсы хранилища данных формируются на основе фиксируемых на протяжении продолжительного периода времени моментальных снимков баз данных оперативной информационной системы и, возможно, различных внешних источников. В хранилищах данных применяются технологии ба данных , OLAP, глубинного анализа данных, виртуализации данных.

Основные характеристики хранилищ данных:

Содержит исторические данные;

Хранит подробные сведения, а также частично и полностью обобщенные данные;

Данные в основном являются статическими;

Средняя и низкая интенсивность обработки транзакций;

Непредсказуемый способ использования данных;

Предназначено для проведения анализа;

Ориентировано на предметные области;

Поддержка принятия стратегических решений;

Обслуживает относительно малое количество работников руководящего звена.

Термин OLAP служит для описания модели представления данных и соответственно технологии их обработки в хранилищах данных. В OLAP применяется многомерное представление агрегированных данных для обеспечения быстрого доступа к стратегически важной информации в целях углубленного анализа. Приложения OLAP должны обладать следующими основными свойствами:

Многомерное представление данных;

Поддержка сложных расчетов;

Правильный учет фактора времени.

Преимущества OLAP:

Повышение производительности производственного персонала, разработчиков прикладных программ. Своевременный доступ к стратегической информации;

Предоставление пользователям достаточных возможностей для внесения собственных изменений в схему;

Приложения OLAP опираются на хранилища данных и системы OLTP, получая от них актуальные данные, что дает сохранение контроля целостности корпоративных

данных;

Уменьшение нагрузки на системы OLTP и хранилища данных.

Основные элементы и операции OLAP

В основе OLAP лежит понятие гиперкуба, или многомерного куба данных, в ячейках которого хранятся анализируемые данные.

Факт- это числовая величина, которая располагается в ячейках гиперкуба.

Измерение-это множество объектов одного или нескольких типов, организованных в виде иерархической структуры и обеспечивающих информационный контекст числового показателя. Измерение принято визуализировать в виде ребра многомерного куба. Объекты, совокупность которых образует измерение, называются членами измерений. Члены измерений визуализируют как точки или участки, откладываемые на осях гиперкуба.

Ячейка- атомарная структура куба, соответствующая полному набору конкретных значений измерений.

Иерархия-группировка объектов одного измерения в объекты более высокого уровня. Иерархии в измерениях необходимы для возможности агрегации и детализации значений показателей согласно их иерархической структуре. Иерархия целиком основывается на одном измерении и формируется из уровней.

Заключение

Организация распределенной базы необходима для компаний, осуществляющих различные виды деятельности, если в их повседневной работе возникает потребность решения следующих задач:

Необходимость оперативного получения информации из баз данных дистанционно отдаленных подразделений;

Необходимость ограничения и осуществления контроля изменения данных в дистанционно отдаленных подразделениях компании;

Необходимость введения централизованного изменения структуры и правил работы баз данных для работы всех дистанционно отдаленных подразделений и

юридических лиц.

Основная задача систем управления распределенными базами данных состоит в обеспечении 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.

Таким образом, распределенные информационные системы являются неотъемлемой частью современной информационной системы. При этом должны обеспечиваться: простота использования системы; возможности автономного функционирования при нарушениях связности сети или при административных потребностях; высокая степень эффективности.