

**Поляков Евгений Альбертович**

Магистрант НАЧОУ ВПО СГА

**Направление:** Информатика и вычислительная техника

**Магистерская программа:**

Распределенные автоматизированные системы

### **Особенности моделирования распределенных систем хранения данных**

**Аннотация:** В статье рассматривается роль современных распределенных систем хранения данных, а также особенности и способы их моделирования.

**Ключевые слова:** Распределенная база данных, система управления распределенной базой данных, математическая модель распределённой реплицированной базы данных

В современном мире все большую роль играют технологии, обеспечивающие эффективное хранение и обработку данных. Связано это с наблюдаемым с конца прошлого века лавинообразным ростом информации. Современные задачи и приложения, связанные с анализом данных, предъявляют особые требования к вычислительным ресурсам, значительно превышающие возможности отдельных компьютеров.

Бурное развитие сетевых технологий последних десятилетий требует постоянного повышения уровня организации коммуникации. Привнести упорядоченность в коммуникационные структуры призваны распределенные системы, управляющие множеством объединенных в сеть компьютеров, представляя их единым целым. При этом предназначение распределенных систем может широко варьироваться: от манипулирования информационными потоками между отдаленными уголками планеты, до контроля кластеров относительно недалеко разнесенных машин.

В настоящее время причина появления систем распределенных вычислений заключается в том, что в крупных автоматизированных информационных системах, построенных на основе корпоративных сетей, не всегда удается организовать централизованное размещение всех баз данных и СУБД на одном узле сети. В связи с этим системы распределенных вычислений тесно связаны с системами управления распределенными базами данных.

Распределенная база данных (РБД) представляет собой совокупность логически взаимосвязанных баз данных, распределенных в компьютерной сети [1].

Система управления распределенной базой данных – это программная система, которая обеспечивает управление распределенной базой данных и прозрачность ее распределенности для пользователей.

Распределенная база данных может объединять базы данных, поддерживающие любые модели (иерархические, сетевые, реляционные и объектно-ориентированные базы данных) в рамках единой глобальной схемы. Подобная конфигурация должна обеспечивать для всех приложений прозрачный доступ к любым данным независимо от их местоположения и формата [1].

РБД состоит из набора узлов, связанных коммуникационной сетью, в которой каждый узел – это полноценная СУБД сама по себе, а узлы взаимодействуют между собой таким образом, что пользователь любого из них может получить доступ к любым данным в сети так, как будто они находятся на его собственном узле.

К основным принципам создания и функционирования распределенных баз данных относятся следующие [3]:

1. Локальная независимость. Узлы в распределённой системе должны быть независимы, или автономны, все операции на узле контролируются этим узлом.

2. Отсутствие опоры на центральный узел. Локальная независимость предполагает, что все узлы в распределённой системе должны рассматриваться как равные.

3. Непрерывное функционирование.

4. Независимость от расположения. Пользователи не должны знать, где именно данные хранятся физически и должны поступать так, как если бы все данные хранились на их собственном локальном узле.

5. Обработка распределённых запросов. Суть в том, что для запроса может потребоваться обращение к нескольким узлам. В такой системе может быть много возможных способов пересылки данных, позволяющих выполнить рассматриваемый запрос.

6. Независимость от операционной системы. Возможность функционирования СУБД под различными операционными системами.

7. Независимость от сети. Возможность поддерживать много принципиально различных узлов, отличающихся оборудованием и операционными системами, а также ряд типов различных коммуникационных сетей.

Необходимо отметить, что практическая реализация распределённых вычислений осуществляется через отступление от некоторых рассмотренных выше принципов создания и функционирования распределённых систем. В зависимости от того, какой принцип приносится в «жертву» (отсутствие центральной установки, непрерывность функционирования, согласованного состояния данных и др.) можно выделить несколько самостоятельных направлений в технологиях распределённых систем – технологии «Клиент-сервер», технологии реплицирования, технологии объектного связывания.

Проектирование и организация распределённой баз данных является трудоемким и дорогостоящим процессом. Чаще всего при проектировании необходимо заранее выполнить расчет характеристик распределённой базы данных. Предварительное моделирование также является необходимым, так как большинство реальных систем должны работать в режиме 24/7/365, и простой

таких систем вследствие технических проблем приводят к большим финансовым потерям [5].

Математическая модель распределенной реплицированной базы данных представляет собой комплекс моделей ее основных компонентов. Узлы системы (базы данных) – моделируются системами массового обслуживания различных видов с различными параметрами. Обычно варьируется наиболее важный параметр – распределение поступления заявок, выбор распределения зависит от конкретной задачи. Не менее важным компонентом модели является модель связи между узлами, обычно она также моделируется СМО, которая имитирует задержку в сети и работу основных ее элементов. Также в большинстве моделей распределенных систем учитывается модель репликации, качество репликации (качество выбор данных для реплицирования), модели доступа к данным (модель локального доступа, Hot-Spot, смешанные модели), контроль совместного доступа к данным, модели различных типов транзакций (таблица 1.1) [4].

Таблица 1.1

**Возможные способы моделирования отдельных компонентов модели**

Компоненты модели	Способ моделирования			
	М/М/м	Сеть из m M/M/1	Сеть из m M/G/1	Каждый узел отдельная сеть (сеть из сетей)
Узлы распределенной БД	М/М/м	Сеть из m M/M/1	Сеть из m M/G/1	Каждый узел отдельная сеть (сеть из сетей)
Моделирование влияния сети	Постоянное время отклика, неограниченная емкость (M/D/inf)	Непостоянное время отклика, неограниченная емкость (M/D/inf)	Экспоненциальная зависимость для времени отклика, модель с ограниченным накопителем (M/M/1)	Общее время отклика, модель с ограниченным накопителем (M/G/1)
Модель репликации	нет	Полная репликация	Частичная 1 измерение	Частичная 2 измерения
Качество репликации	Не используется	Моделируется качество выбора местонахождения копии	Моделируется качество выбора информации для реплицирования	Моделируется качество выбора местонахождения копии и выбора информации для

				реплицирования
Доступ к данным	Однородный (Доступ к каждому объекту данных равновероятен)	Locality (Доступ к локальным данным более вероятен)	Hot-spot (Доступ к одной группе данных более вероятен)	Locality + Hot-spot
Модель обработки транзакций	Только запросы на выборку (только чтение)	Только запросы на изменение (только запись)	Изменение + Выборка	Более 2 типов транзакций
Разрешение конфликтов	Игнорируется	2PL (2-фазные блокировки). Полные блокировки.	2PL, блокировки чтения и записи	Другие

Кроме того, необходимо отметить, что при построении модели распределенной БД, очень важным является выбор критерия, по которому в дальнейшем будет оцениваться производительность работы системы. Наиболее часто используемыми критериями являются – время обработки транзакции и пропускная способность системы. Данные критерии являются внешними, так как они непосредственно влияют на удобство пользования системой конечными пользователями. Внутренними критериями быстродействия являются – среднее количество узлов, доступ к которым был произведён из одной транзакции, вероятность возникновения конфликтов при доступе к данным, количество и частота не выполненных транзакций, количество передаваемых сообщений внутри системы, доступность данных (вероятность того, что операция ввода/вывода может получить доступ к требуемым данным) [1].

На основании этого, можно сказать, что распределенная система характеризуется распределенной структурой, т.е. набором дискретных (отделенных друг от друга) элементов или подсистем и связями между этими элементами. В соответствии с распределением элементов распределены и функции (алгоритмы работы) системы. Распределенные элементы обмениваются сообщениями (сигналами), также распределенными в пространстве [2].

Моделирование применяется в тех случаях, когда проведение реального эксперимента сопряжено с опасностью, высокими экономическими и временными затратами или неудобен в масштабе пространства и времени.

Также моделирование СХД позволяет оценить целевые параметры системы исходя из ее архитектуры и технических характеристик низкоуровневых элементов.

### Литература

1. Цилькер Б.Я., Орлов С.А. Организация ЭВМ и систем. СПб.: Питер, 2012.
2. Бройдо В.Л. Вычислительные системы, сети и телекоммуникации. СПб.: Питер, 2010.
3. Резников Б.В. Особенности организации системы управления базами данных хранения многоуровневых структурных моделей // Известия Южного Федерального университета. Технические науки. 2009. № 6.
4. Коровин Е.Н. Методология прогнозирования и оптимального управления территориально распределенными социально-экономическими системами на основе трансформации информации и многовариантного моделирования. Воронеж: РГБ ОД, 2010.
5. Олзоева С.И. Моделирование и расчет распределенных информационных систем. Учебное пособие. Улан-Удэ: Изд-во ВСГТУ, 2008.

© Бюллетень магистранта 2014 год № 6