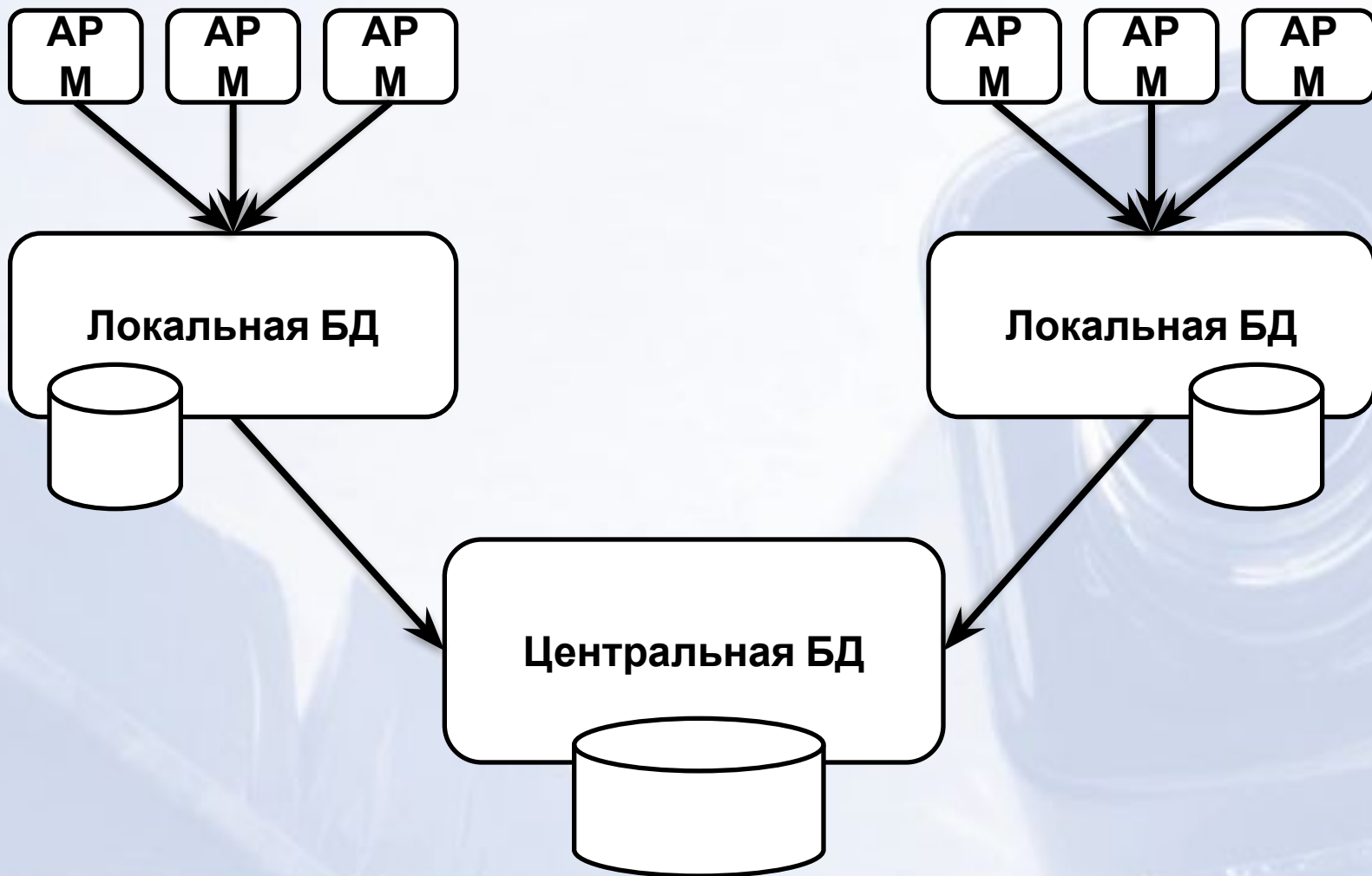


Лекция №6


Управление параллелизмом

каф. КИБЭВС
И.В. Горбунов

Под распределенной (Distributed DataBase - DDB)
обычно понимают базу данных,
декомпозированную и фрагментированную на
несколько узлов вычислительной сети, с
возможным управлением различными СУБД.



Распределенная БД должна обладать:

1. Локальными и глобальными (распределенными) средствами доступа к данным(СУБД).
 2. Единообразной логикой прикладных программ во всех АРМах сети.
 3. Малым временем реакции на запросы пользователей.
 4. Надежностью, исключающей разрушения целостности системы в случае выхода из строя ее отдельных компонент(узлов).
 5. Открытостью, позволяющей наращивать объем локальных БД и добавлять новые АРМ.
- 

Распределенная БД должна обладать:

6. Развитой системой управления резервным копированием и восстановления данных на случай сбоев.
7. Защищенностью, следящей за соблюдением привилегий доступа к данным.
8. Высокой эффективностью, за счет выбора оптимальных алгоритмов использования сетевых ресурсов.
9. Развитым репликационным механизмом, позволяющим размещать обновленные копии данных в сети оптимальным образом.

Принципы построения распределенной БД:

1. Минимизация интенсивности обмена данными (сетевое трафика).
2. Оптимальным размещением серверных и клиентских приложений в сети.
3. Декомпозиция данных на часто и редко используемые сегменты (для правильной настройки репликации - размещение наиболее часто используемых данных на АРМ конечных пользователей).
4. Периодическое сохранение копий данных и выполнение действий по поддержке целостности распределенной информационной системы.

Критерии построения распределенной БД:

1. Всесторонний анализ информационных потребностей предметной области с выявлением объемов хранимых данных, их сложности, достоверности, взаимосвязанности.
2. Моделирование предполагаемого сетевого трафика при работе распределенной БД с различными моделями репликации данных.
3. Кластеризация элементов данных и программ их обработки. Цель - добиться максимальной автономности и слабосвязанности кластеров.
4. Привязка кластеров данных к вероятным пользователям или АРМ.
5. Поддержка эталонной копии данных и ограничение репликационного механизма.
6. Разработка и реализация правил приведения локальных и центральной БД в непротиворечивое состояние.

Распределенные архитектуры БД принято подразделять по типам на

- Системы не дублирующего 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)

Достаточные свойства для корректной работы распределенной БД

Локальная автономия

Это качество означает, что управление данными на каждом из узлов распределенной системы выполняется локально. База данных, расположенная на одном из узлов, является неотъемлемым компонентом распределенной системы. Будучи фрагментом общего пространства данных, она, в то же 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 (известное "24 часа в сутки, семь дней в неделю") в рамках DDB (Distributed DataBase) 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 (то есть на различных узлах) размещения данных, логически представляющих собой единое целое. Существует фрагментация двух типов: горизонтальная и вертикальная. Первая означает хранение строк одной таблицы на различных узлах (фактически, хранение строк одной логической таблицы в нескольких идентичных физических таблицах на различных узлах). Вторая означает распределение столбцов логической таблицы по нескольким узлам.

Прозрачное тиражирование

Тиражирование данных - это асинхронный (в общем случае) 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.

Обработка распределенных транзакций

Это качество DDB можно трактовать как возможность выполнения операций обновления распределенной базы данных (INSERT, UPDATE, DELETE), не разрушающее целостность и согласованность данных. Эта цель достигается применением двухфазового или двухфазного протокола фиксации транзакций (two-phase commit protocol), ставшего фактическим стандартом обработки распределенных транзакций. Его применение гарантирует согласованное изменение данных на нескольких узлах в рамках распределенной (или, как ее еще называют, глобальной) транзакции.

Прозрачность сети

Доступ к любым базам данных может осуществляться по сети. Спектр поддерживаемых конкретной СУБД сетевых протоколов не должен быть ограничением системы с распределенными базами данных. Данное качество формулируется максимально широко - в распределенной системе возможны любые сетевые протоколы.

Независимость от оборудования

Это свойство означает, что в качестве узлов распределенной системы могут выступать компьютеры любых моделей и производителей - от мэйнфреймов до "персоналок".

Независимость от операционных систем

Это качество вытекает из предыдущего и означает многообразие операционных систем, управляющих узлами распределенной системы.

Независимость от систем управления

Это качество означает, что в распределенной системе могут мирно сосуществовать СУБД различных производителей, и возможны операции поиска и обновления в базах данных различных моделей и форматов.

Спасибо за внимание!!!