

Принципы управления распределенной информацией

План

- Распределенные базы данных
- Характеристики РаБД
- Желаемый сценарий управления распределенной информацией
- Модели РаБД
- Проблемы управления распределенной информацией



Распределенные базы данных

Распределенная база данных (РаБД) – это совокупность логически взаимосвязанных баз данных, распределенных в компьютерной сети.

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



Характеристики РаБД (12 правил)

- **Локальная автономность.** Локальные данные должны находится под локальным владением и управлением, включая функции безопасности, целостности, представления данных в памяти. Исключением может быть ситуация, когда ограничения целостности охватывают данные нескольких узлов или когда управление распределенной транзакцией осуществляется некоторым внешним узлом.
- **Никакой конкретный сервис не должен возлагаться на какой либо специально выделенный центральный узел.** Соблюдение этого правила, то есть принципа децентрализованности функций РаСУБД, позволяет избежать узких мест.



Характеристики РаБД (12 правил)

- **Непрерывность функционирования.** Система не должна останавливаться в случае необходимости добавления нового узла или удаления в распределенной среде некоторых данных, изменения определения метаданных и даже (что довольно сложно) осуществления перехода к новой версии СУБД на отдельном узле.
- **Независимость от местоположения.** Пользователи и приложения не обязаны знать о том, где физически располагаются данные.



Характеристики РаБД (12 правил)

- **Независимость от фрагментации.** Фрагменты данных должны поддерживаться и обрабатываться средствами РаСУБД таким образом, чтобы пользователи или приложения могли бы вообще ничего не знать об этом. Более того РаСУБД должна уметь обходить при обработке запросов фрагменты, не имеющие к ним отношения.
- **Независимость от тиражирования.** Те же принципы независимости и прозрачности относятся и к механизму тиражирования.



Характеристики РаБД (12 правил)

- **Распределенная обработка запросов.** Обработка запросов должна производиться распределенным образом.
- **Управление распределенными транзакциями.** На РаБД необходимо распространить механизмы управления транзакциями и управления одновременным доступом.
- **Независимость от оборудования.** Одно и тоже программное обеспечение РаСУБД должно выполняться на различных аппаратных платформах.



Характеристики РаБД (12 правил)

- **Независимость от операционных систем.** Одно и то же программное обеспечение РаСУБД должно выполняться под управлением различных операционных систем.
- **Независимость от сети.** Узлы могут быть связаны между собой с помощью множества разнообразных сетевых и коммуникационных средств.
- **Независимость от СУБД.** В функционировании РаСУБД должны иметь возможность участвовать различные локальные СУБД.



Характеристики РаБД

Крайне желательно было бы иметь системы, удовлетворяющие всем 12 правилам. Но пока пока таких систем нет.

Четыре шага для перехода к управлению РаБД:

- **Удаленный запрос** (один удаленный запрос – одна транзакция)
- **Удаленная единица работы** (группа удаленных запросов на узле как транзакция, транзакция охватывает только один узел)
- **Распределенная единица работы** (каждый запрос относится только к одному узлу, но запросы объединены в транзакцию)
- **Распределенный запрос** (транзакция – группа распределенных запросов, охватывающих несколько узлов)



Желаемый сценарий управления распределенной информацией

- Надстроенная над существующей системой «оболочка управления информацией»
- «Правильные» принципы распределенной обработки
- Возможность разработки новых приложений относительно глобального представления данных
- Новые приложения имеют доступ к существующим данным и к вновь создаваемым объектам
- Возможность применять подходящие модели представления данных для разных компонентов среды



Желаемый сценарий управления распределенной информацией

- Автоматическая оптимизация доступа к распределенным данным
- Расширение традиционных функций СУБД с учетом распределенного характера
- Разумное распределение СУБД по всей среде
- Распределение метаданных
- Обновляемые алгоритмы оценки затрат функционирования



Модели РаБД

Однородные системы



Одинаковые СУБД



Модели РаБД

Неоднородные системы

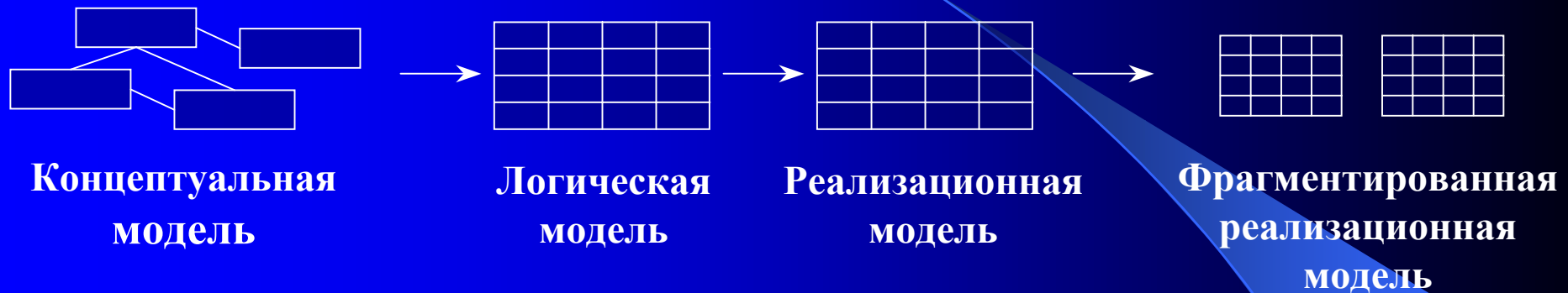


Различные СУБД



Методы построения РаБД «сверху вниз» и «снизу вверх»

Метод «сверху вниз» аналогичен проектированию центральных БД



Метод «снизу вверх» применяется при интеграции унаследованных систем



Фрагментация данных

Фрагментация – декомпозиция объектов базы данных на две и более частей. Поддерживается глобальная схема, позволяющая воссоздать из имеющихся фрагментов логически централизованную таблицу или другую структуру базы данных.

Декомпозиция может строиться по методам зависящим от данных (например, фрагментация по диапазонам) и не зависящим от данных (карусельная фрагментация).



Фрагментация реляционной таблицы

Вертикальная
фрагментация
(по столбцам)

The diagram illustrates vertical fragmentation of a 5x2 table into three separate tables. The first table contains the first two columns of the original table. The second table contains the next two columns. The third table contains the final column.

Горизонтальная
фрагментация
(по строкам)

The diagram illustrates horizontal fragmentation of a 5x2 table into three separate tables. The first table contains the first two rows of the original table. The second table contains the next two rows. The third table contains the final row.

--	--	--	--	--

Тиражирование данных

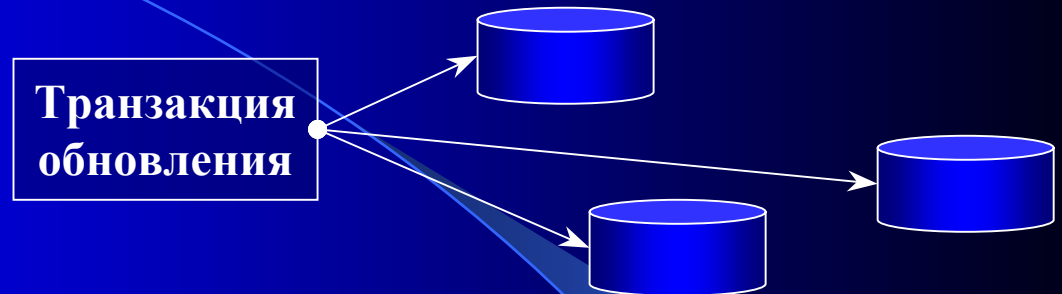
Тиражирование (репликация) – создание дубликатов данных.

Репликатy – это множество различных физических копий некоторого объекта базы данных, для которых в соответствии с определенными правилами поддерживается синхронизация (идентичность) с некоторой «главной» копией.

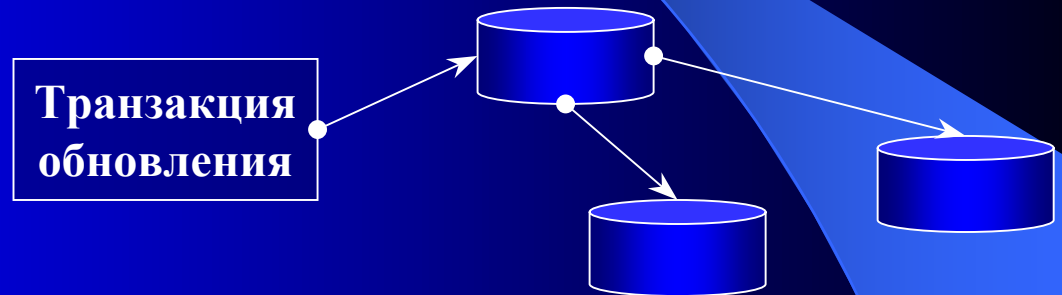


Модели тиражирования данных

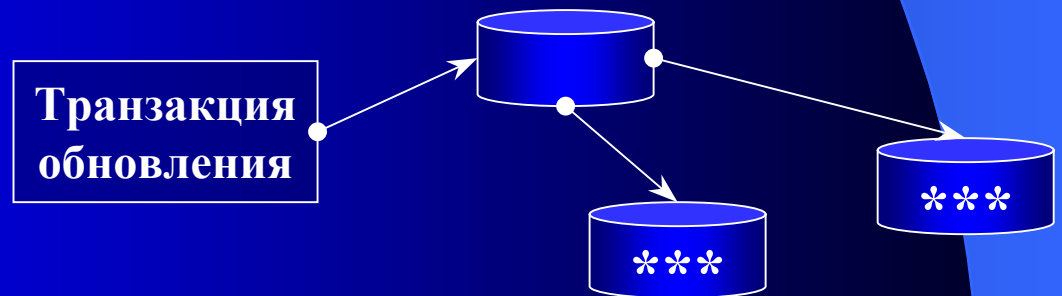
Одновременное обновление
(с управлением параллелизмом)



Распространение обновления



Запланированная синхронизация дубликатов ТОЛЬКО для чтения



Проблемы управления распределенной информацией

- **Нарушение семантической целостности.** Несоответствие между представлениями и форматами данных, поступающих из разных источников.
- **Произвольные выборки данных.** Должны быть средства для запроса сведений о структуре БД и процессах консолидации.
- **Безопасность.** В условиях распределенности и неоднородности резко осложняется поддержка безопасности.
- **Масштабируемость по числу узлов.** Алгоритмы, работающие для небольшого числа узлов, не масштабируются на среду из тысяч узлов.
- **Обработка транзакций.** Локальные СУБД распределенной среды могут иметь различные модели транзакций.
- **Проектирование распределенных сред.** Пока не формализованы методы разработки распределенных сред, а старые методы уже не применимы.
- **Распределенная обработка запросов.** Глобальная оптимизация должна учитывать индивидуальные характеристики СУБД. При динамических изменениях необходимо заново производить оптимизацию.





Топорец Александр Юрьевич
email: 4sale@mail.ru
www: www.stoporets.narod.ru

