



Технология реплицирования данных

Основная идея реплицирования



Основная идея реплицирования заключается в том, что пользователи работают автономно с общими данными, растиражированными по локальным БД, обеспечивая с учетом отсутствия необходимости передачи и обмена данными по сети **максимальную производительность для своих вычислительных установок.**

Репликой называют особую копию БД для размещения на другом компьютере сети с целью автономной работы пользователей с одинаковыми (согласованными) данными общего пользования.



Программное обеспечение СУБД, использующих технологию реплицирования, дополняется функциями тиражирования (реплицирования) баз данных, причем тиражироваться могут как сами данные и их структуры, так и системный каталог с информацией о размещении реплик. При этом возникают две проблемы в реализации одного из основных принципов построения и функционирования распределенных систем, а именно - непрерывности согласованного состояния данных:

- обеспечение согласованного состояния во всех репликах количества и значений общих данных;
- обеспечение согласованного состояния во всех репликах структуры данных.



1. Обеспечение согласованного состояния общих данных



Основывается на реализации одного из двух принципов:

- принципа непрерывного размножения обновлений (любое обновление данных в любой реплике должно быть немедленно размножено);
- принципа отложенных обновлений (обновление реплик должно быть отложено до специальной команды или ситуации).



Принцип непрерывного размножения обновлений



Принцип непрерывного размножения обновлений является основным при построении систем реального времени, где требуется непрерывное и точное соответствие реплик во всех узлах и компонентах системы.

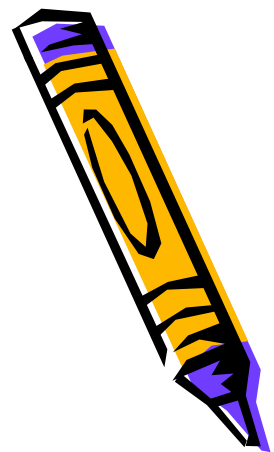
Этот принцип заключается в том, что любая транзакция считается успешно завершённой, если она успешно завершена во всех репликах системы.

Для обнаружения тупиков в реплицированных системах строится и поддерживается специальный граф ожидания транзакций и применяются алгоритмы распознавания тупиков, основанные на технике приоритетов.



Принцип отложенных обновлений

Накопленные в какой-либо реплике изменения данных специальной командой направляются к остальным репликам системы. Возможность конфликтов и тупиков при синхронизации реплик существенно снижается, а если и возникают, то их легко можно разрешить организационно.

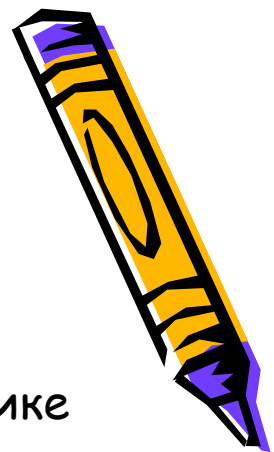


2. Согласованность структуры данных

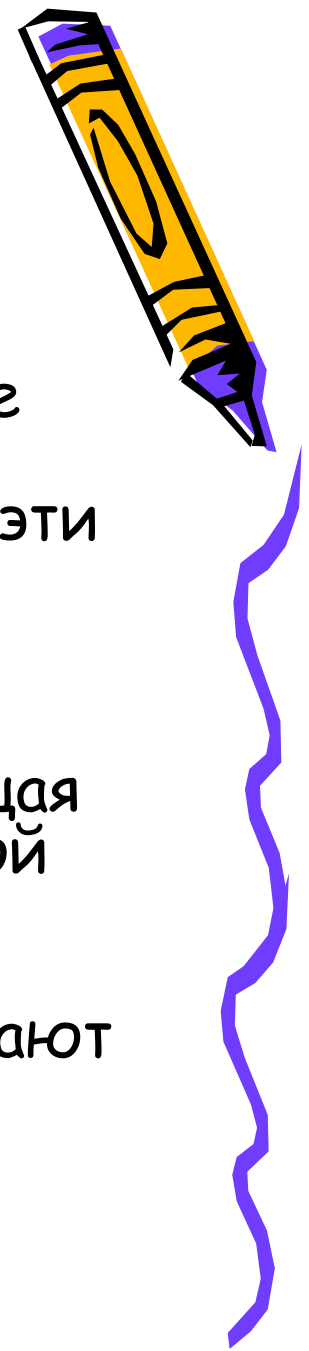
Решение проблемы согласованности структуры данных, осуществляется через частичное отступление от принципа отсутствия центральной установки и основывается на технике «главной» реплики.

Суть этой техники заключается в том, что одна из реплик БД системы является главной. Изменять структуру данных можно только в главной реплике. Эти изменения тиражируются на основе принципа отложенных обновлений через специальную синхронизацию реплик.

Частичность отступления от принципа отсутствия центральной установки заключается в том, что выход из строя главной реплики не влечет гибель сразу всей распределенной системы, т.к. остальные реплики продолжают функционировать автономно. Более того, в таких СУБД администратор системы может преобразовать любую реплику в главную и восстановить работоспособность всей системы.



Частичные реплики

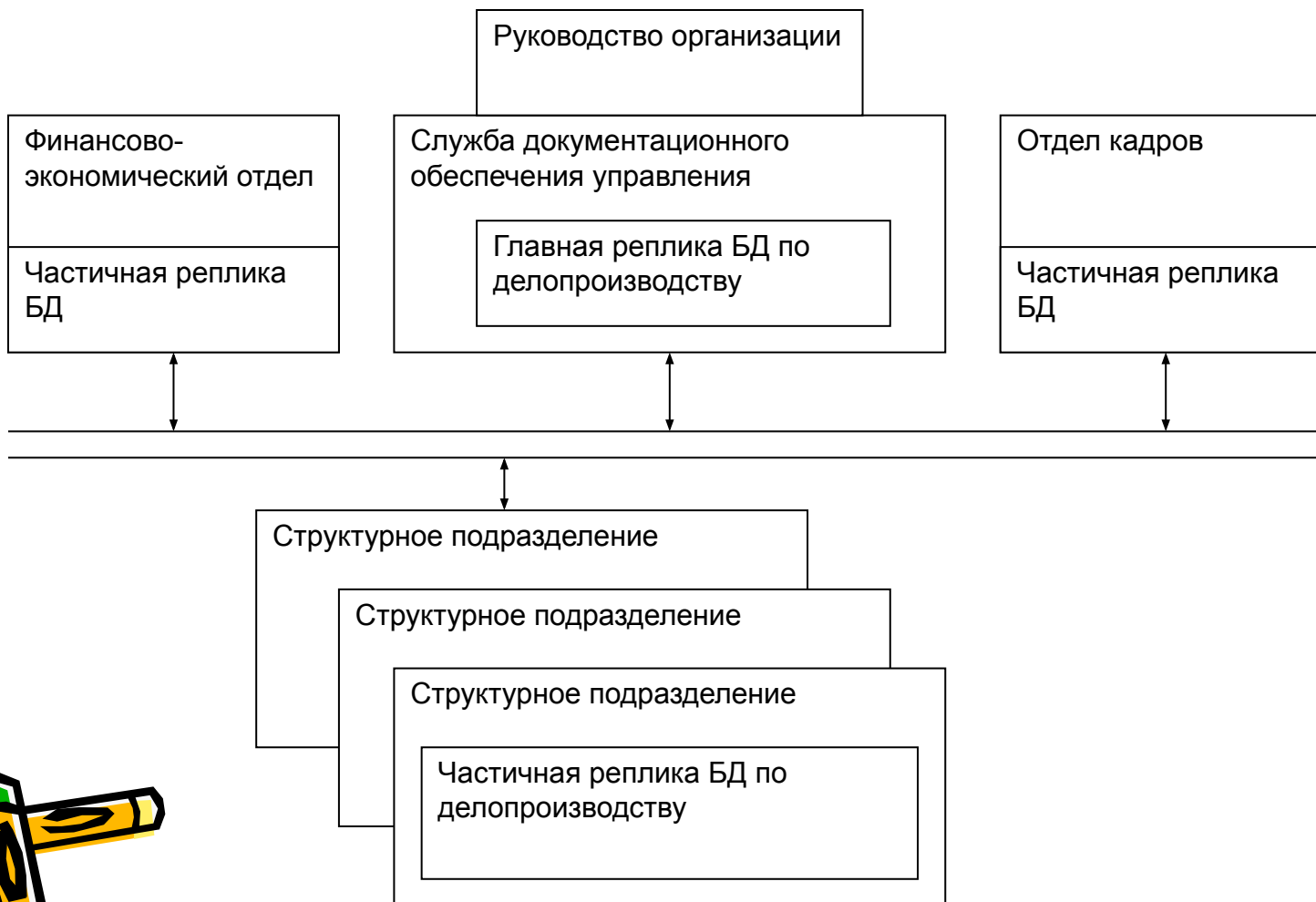
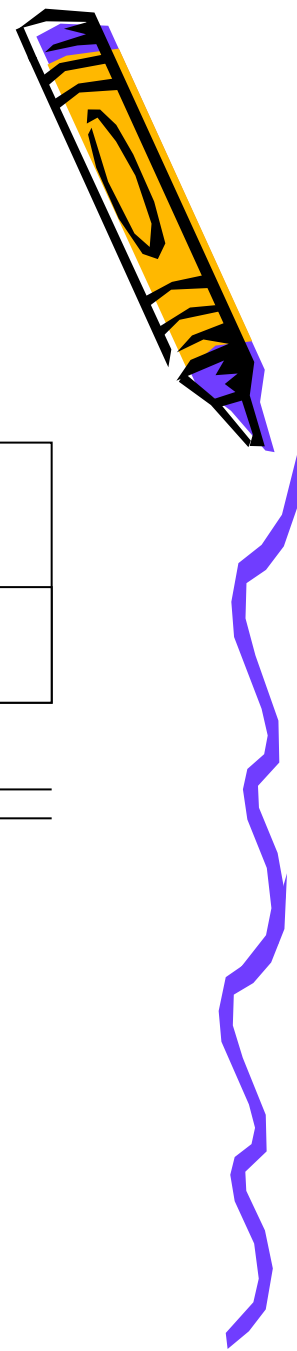


Важным с точки зрения гибкости и эффективности функционирования распределенных ИС, построенных на основе технологий реплицирования, является создание частичных реплик и включение в эти реплики как реплицируемых, так и нереплицируемых данных.

Частичной репликой называется БД, содержащая ограниченное подмножество записей полной реплики. Частичные реплики позволяют решить некоторые проблемы, связанные с разграничением доступа к данным и повышают производительность обработки данных.



Общая схема распределенной ИС по делопроизводству некоторой организации на основе технологий репликации данных.



Заключение

Технологии репликации данных являются экономичным решением по сравнению с дорогостоящими тяжелыми Клиент-серверными системами в тех случаях, когда не требуется обеспечивать большие потоки в сети.

На практике для совместной обработки данных применяются смешанные технологии, включающие элементы объектного связывания данных, репликаций и клиент-серверных решений.

При этом дополнительно к проблеме логического проектирования добавляется не менее сложная проблема транспортно-технологического проектирования информационных потоков, разграничения доступа и т.д., для которых теория, методология. И инструментальные подходы пока не проработаны.

