

Администрирование многопользовательских систем управления баз данных

ФИО преподавателя: Смирнов Михаил Вячеславович

e-mail: smirnovmgupi@gmail.com

Лекция 4

Репликация в многopользовательских БД.

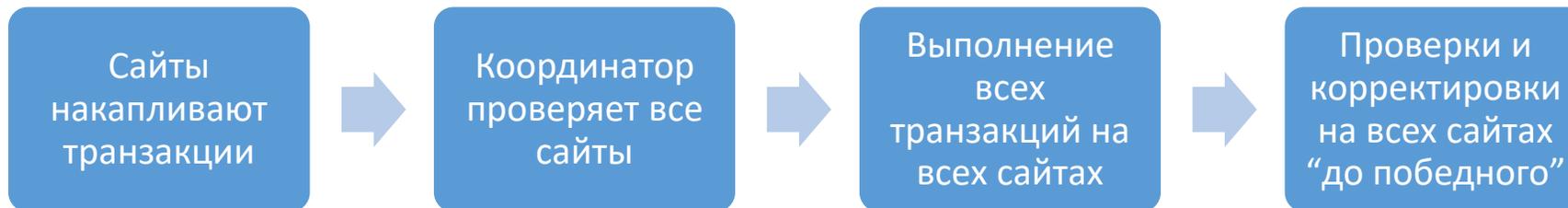
Распределенные данные

- **Распределённая база данных (DDB)** - база данных, составные части которой размещаются в различных узлах компьютерной сети в соответствии с каким-либо критерием.
- Данные представляют собой DDB, только если они связаны в соответствии с некоторым структурным формализмом, реляционной моделью, а доступ к ним обеспечивается единым высокоуровневым интерфейсом.
- Центральная идея – доступность данных в любое время, в любом месте.

Преимущества использования распределенных данных

- Данные доступны людям, которым они нужны, и когда они нужны.
- Система позволяет локальному пользователю автономно оперировать данными (нет проблем с параллельной обработкой).
- Система сокращает сетевой трафик.
- Процесс обеспечения непрерывности бизнеса с точки зрения БД дешевле.

Распределенная транзакция



Чем больше сайтов-участников, тем меньше шансов обойтись "малой кровью" при исполнении общей транзакции.

Общие определения репликации МБД

В общем виде *репликация* – это тиражирование изменений данных с главного сервера БД на одном или нескольких зависимых серверах.

В процессе репликации главный сервер называют *мастер*, а зависимые – *реплики*.

CRUD процедуры выполняются только на *мастере*!

Сервер реплика предоставляет возможность выполнения операции SELECT.

Репликация осуществляется за счет выполнения на репликах инструкций из *бинарных логов* (binlog), располагаемых на мастере.

Измененные данные с мастера не передаются!

Причины репликации МБД

1. *Производительность и масштабируемость.* Снятие с сервера мастера части нагрузки по обработке данных. Чем больше операций чтения приходится на одну операцию записи, тем больше выгоды от репликации.
2. *Отказоустойчивость.* В случае отказа реплики, обязанности по обработке операций чтения перехватывает на себя сервер мастер, тогда как реплика в случае отказа мастера может перехватить на себя функции записи.
3. *Резервирование данных.* Реплику можно остановить и снять dump в любое время.
4. *Отложенные вычисления.* Огромные по объему SQL запросы выгодно выполнять на отдельной реплике.

Термины репликации МБД в SQL Server

Master сервер называется *издателем*.

Информация, которая подвергается репликации называется *публикация*.
Одна публикация – это один согласованный набор данных.

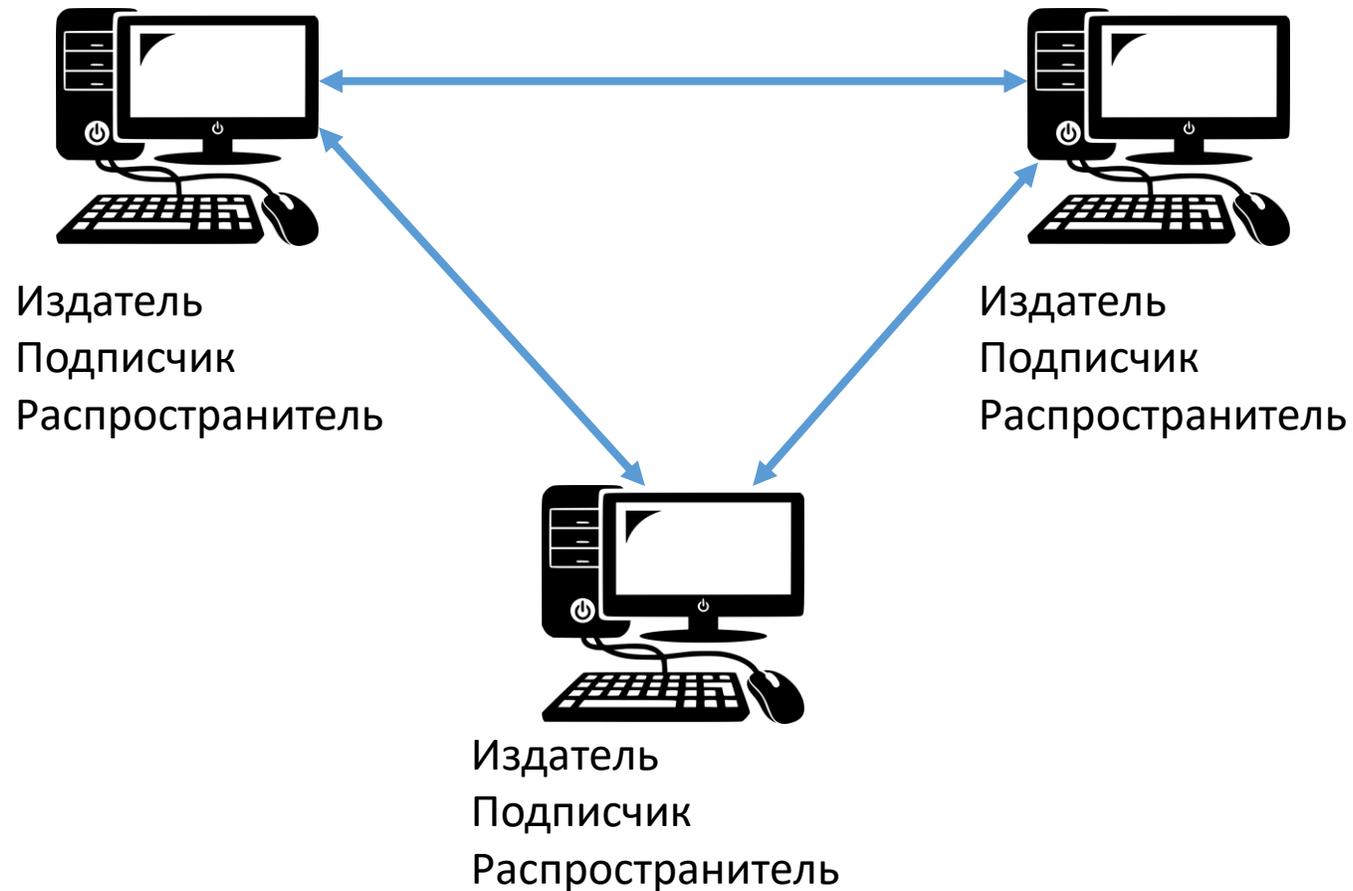
В состав публикации входят *статьи*, которые могут быть:

- целой таблицей или ее частью;
- хранимой процедурой или представлением;
- пользовательской функцией.

Серверы реплики называются *подписчиками*.

Агенты, обеспечивающие постоянную репликацию называются *распространителями*.

Типовая схема репликации



Ограничения публикаций и статей

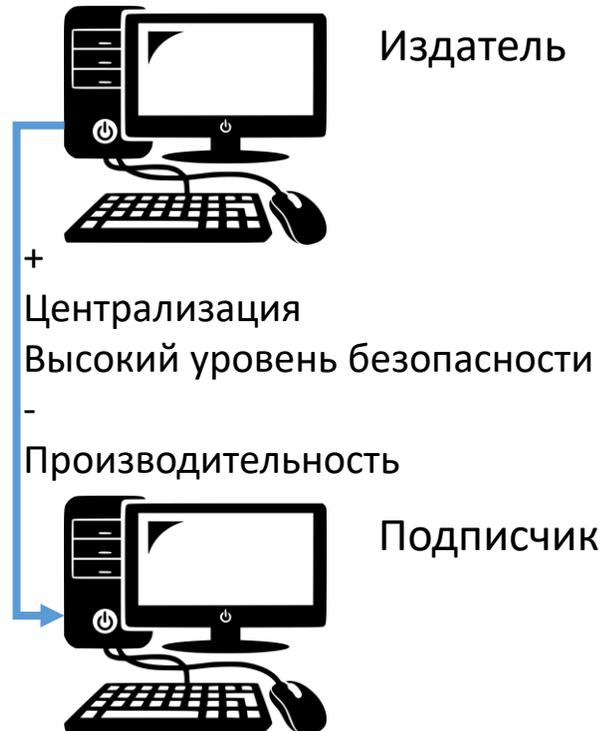
- Статья содержит данные из таблицы и одной или нескольких хранимых процедур
- Таблица может быть как целой, так и подмножеством
- В публикации можно собрать несколько статей
- Каждая публикация должна содержать данные только из одной базы данных
- Подписаться на статью нельзя

Фильтры статей

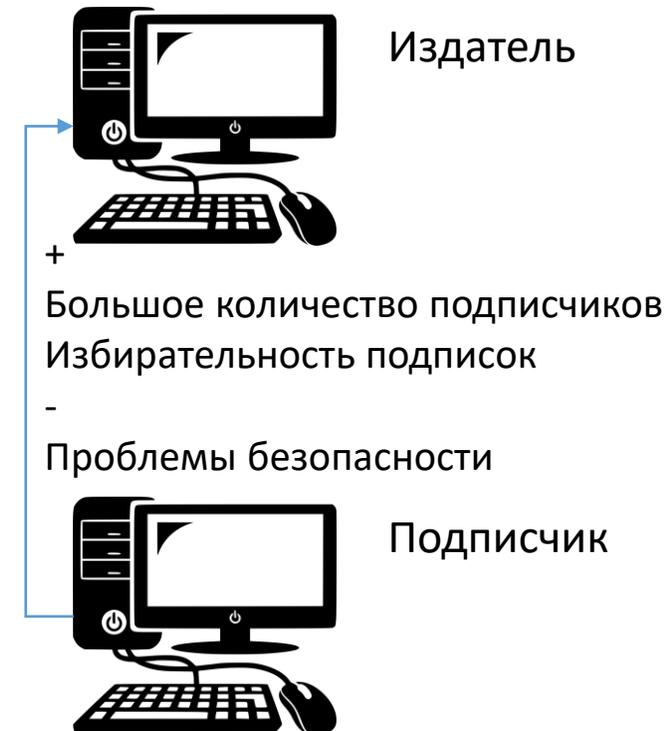


Типы подписок

Принудительная



Запрос



Типы репликации

Аспект	Репликация транзакций	Репликация мгновенного снимка	Репликация слияния	Одноранговая репликация
Транзакционность	Да	Нет	Нет	Да
Сложность	Средняя	Низкая	Средняя	Высокая
Могут ли обновляться подписчики?	Ограничения	Нет	Да	Разделенными данными
Обнаружение конфликта	Нет	Нет	Да	Да
Решение конфликта	Нет	Нет	Да	Нет
Производительность	Высокая	Средняя	Низкая	Высокая

Репликация транзакций



Репликация транзакций

- Distribution – системная база данных Издателя
- В случае репликации через хранимые процедуры, самые дорогие по ресурсам операции будут происходить на Подписчиках
- Перед репликацией транзакций каждый Подписчик должен получить Полную резервную копию БД
- У каждой реплицируемой таблицы должен быть первичный ключ

Репликация мгновенного снимка (snapshot)



Репликация мгновенного снимка (snapshot)

- База данных Distribution не используется
- Файлы передаются Подписчику через определенные интервалы времени
- Только одно направление передачи данных – от Издателя к Подписчику
- Наличие первичного ключа не обязательно

Репликация слияния (merge)



Репликация слияния (merge)

- Данные на сайте собираются Snapshot Agent
- Merge Agent изменяет первоначальную схему данных (добавляет системные таблицы, добавляет первичный ключ, создает триггеры)
- Любое направление передачи данных
- Конфликты обновления решаются методом приоритетов

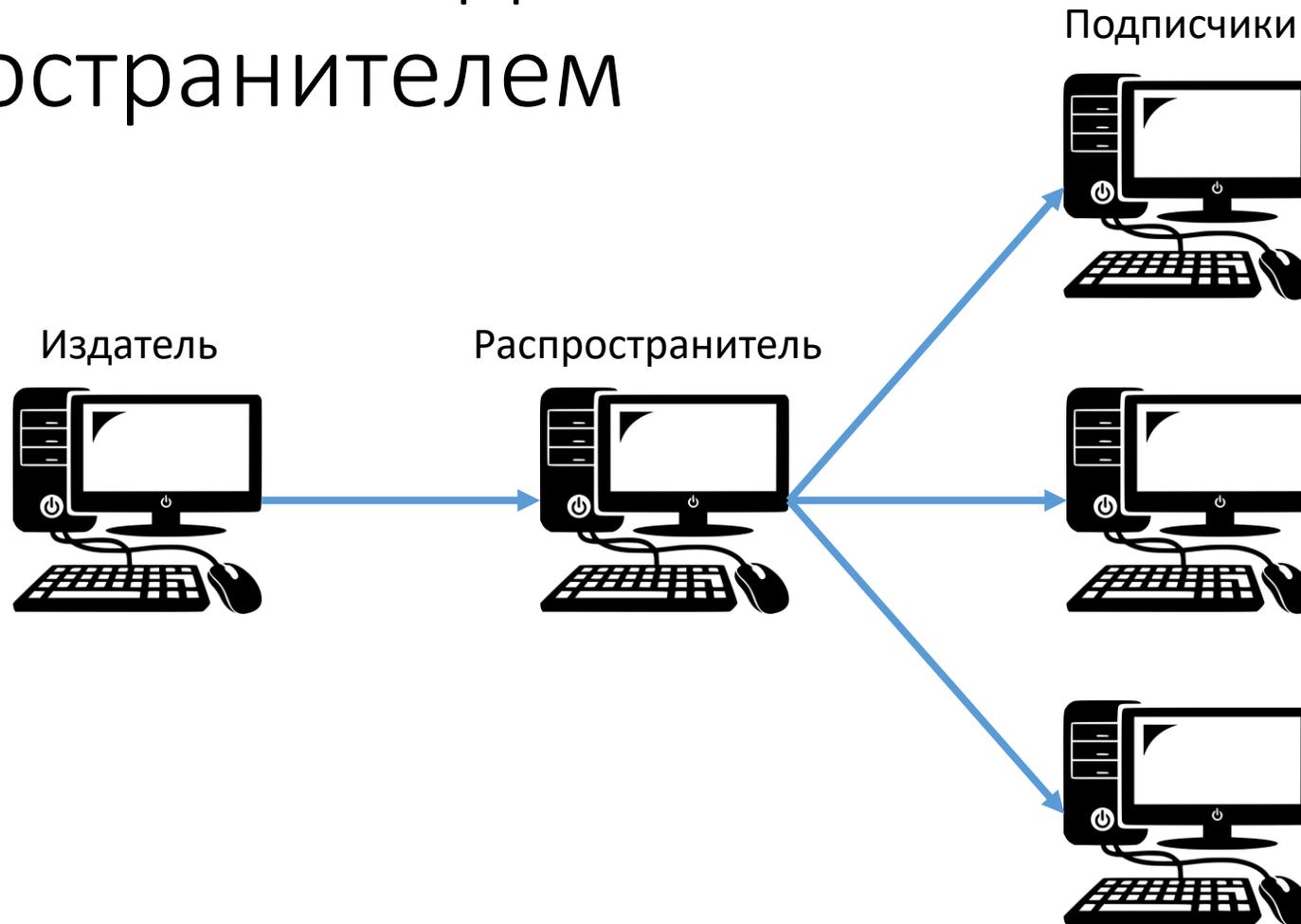
Одноранговая репликация (peer-to-peer)



Одноранговая репликация (peer-to-peer)

- Все серверы являются издателями-распространителями-подписчиками
- Все серверы обладают одним и тем же набором данных
- Каждый сервер манипулирует своим подмножеством данных
- Любой конфликт обновления будет считаться ошибкой

Центральный издатель с распространителем



Чтение на дом

- Петкович, Microsoft SQL Server 2012, стр. 497-517.

Вопросы для самостоятельного изучения

1. Как называется сервер, который содержит полную актуальную копию мастер-сервера БД? В каких случаях используется?
2. Компания имеет филиалы в разных городах. В каждом филиале есть свой сервер БД, который имеет такой же набор данных, что и остальные. Какую репликацию данных стоит реализовать в данном случае. Почему?
3. Какой тип репликации подойдет для схемы на слайде 8? Почему?

Спасибо за внимание!