



Для чего нужен и как работает MySQL Group Replication?

Игорь Ильин
MySQL Support

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

1

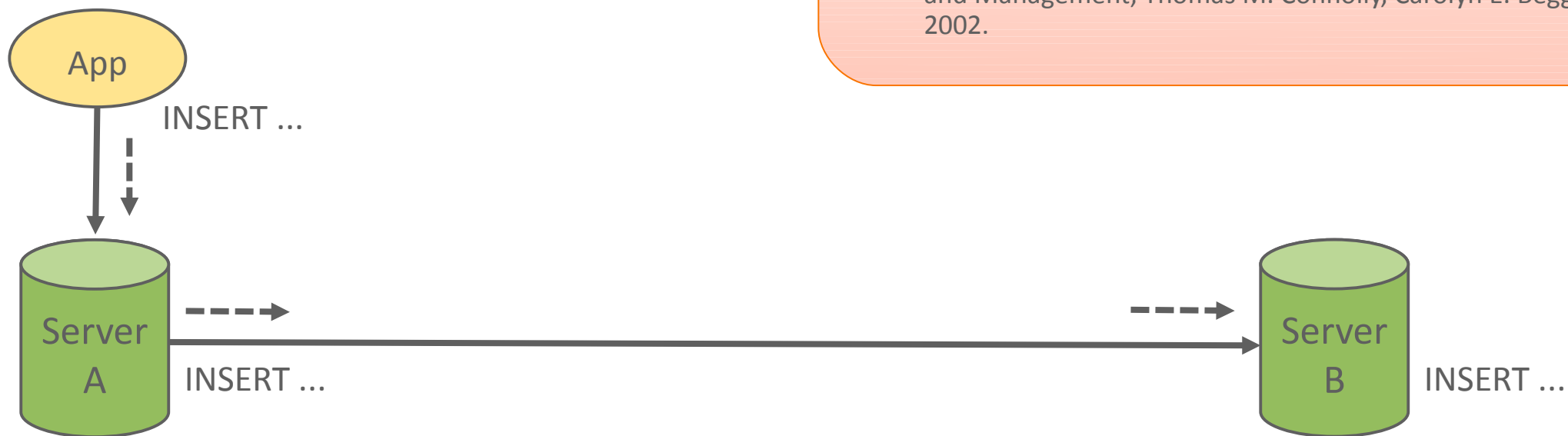
Типы репликации

Репликация баз данных

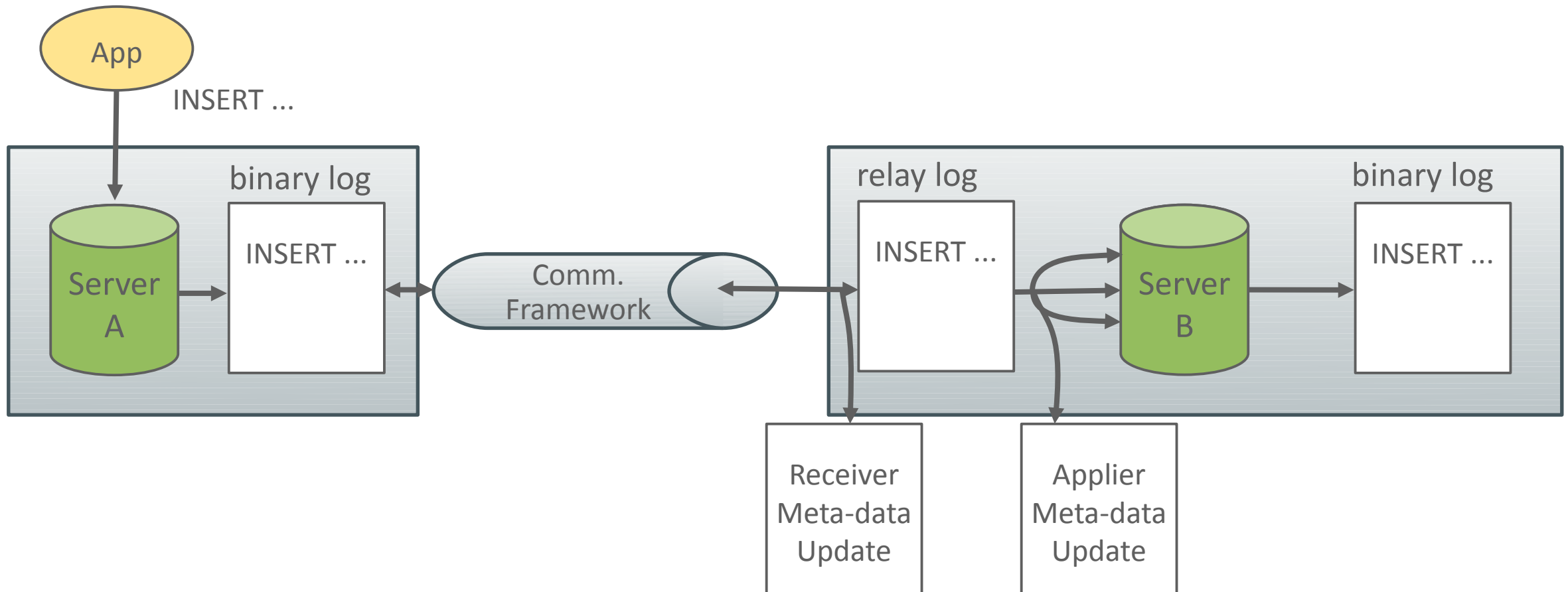
Репликация

“Процесс генерации нескольких копий данных в различных местах”

Database Systems: A Practical Approach to Design, Implementation, and Management, Thomas M. Connolly, Carolyn E. Begg, Third Edition, 2002.



Репликация в MySQL

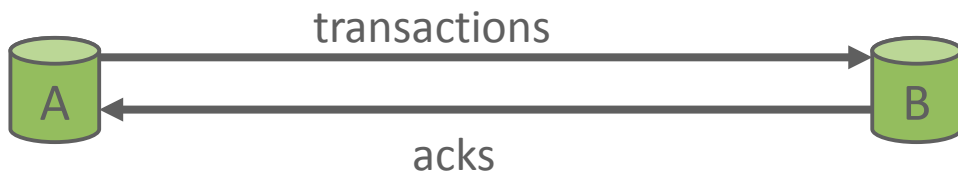


Типы репликации в MySQL



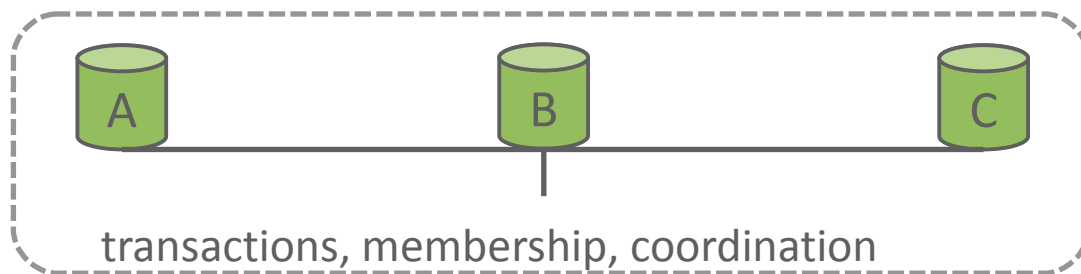
Since 3.23

asynchronous (native)



Since 5.5

semi-synchronous (plugin)



Since 5.7.17

Теперь и в MySQL 8 as of 8.0.1

group replication (plugin)

MySQL Group Replication

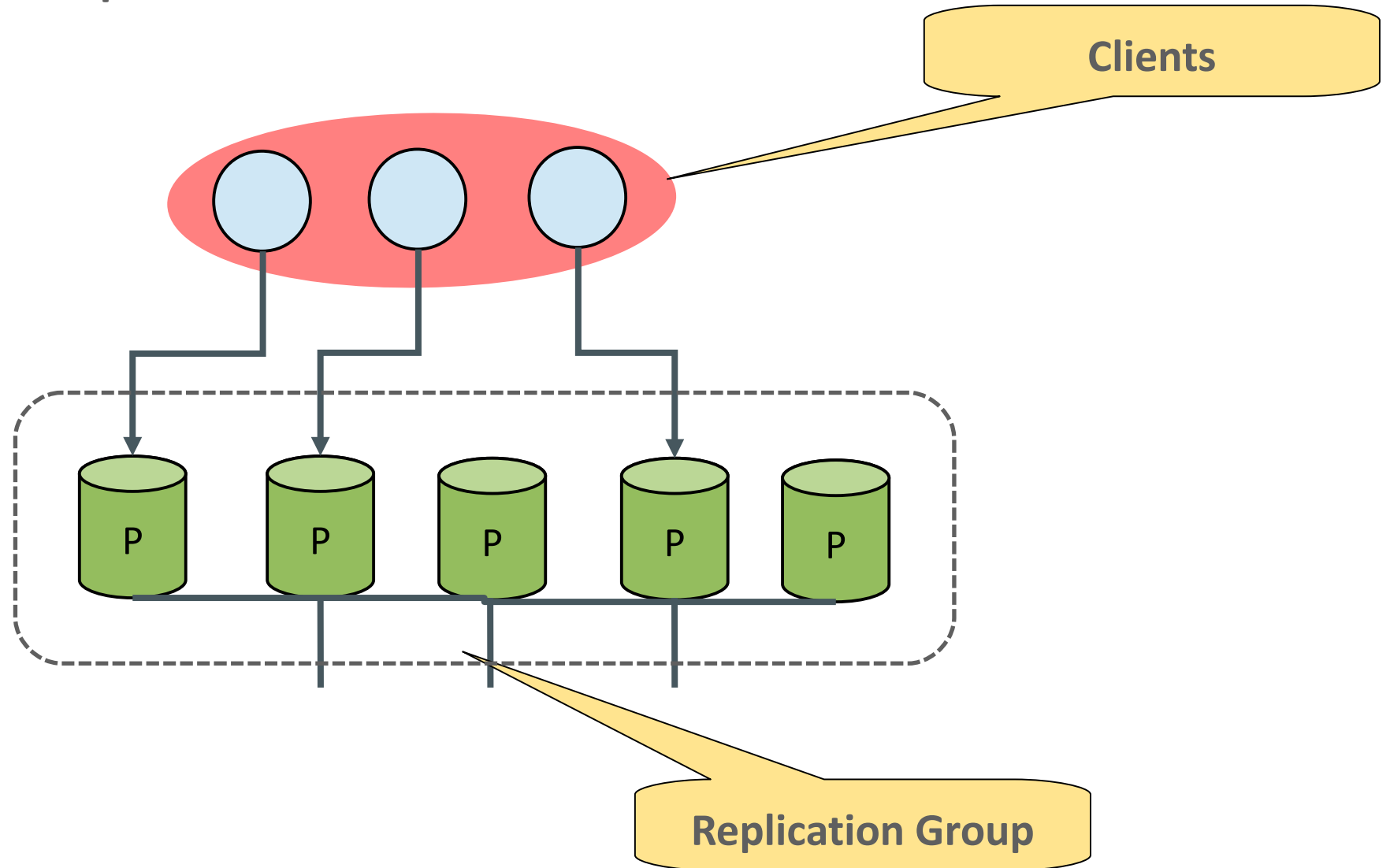
- **Что такое MySQL Group Replication?**

“Плагин для репликации в MySQL, предоставляющий возможность изменения данных на любом из серверов, со встроенной системой автоматического восстановления и с разрешением конфликтов.”

- **Что MySQL Group Replication делает для пользователя?**

- Убирает необходимость ручного переключения в случае отказа одного из серверов.
- Предоставляет отказоустойчивость.
- Позволяет построить систему с возможностью изменять данные на любом сервере.
- Автоматическая реконфигурация.

MySQL Group Replication

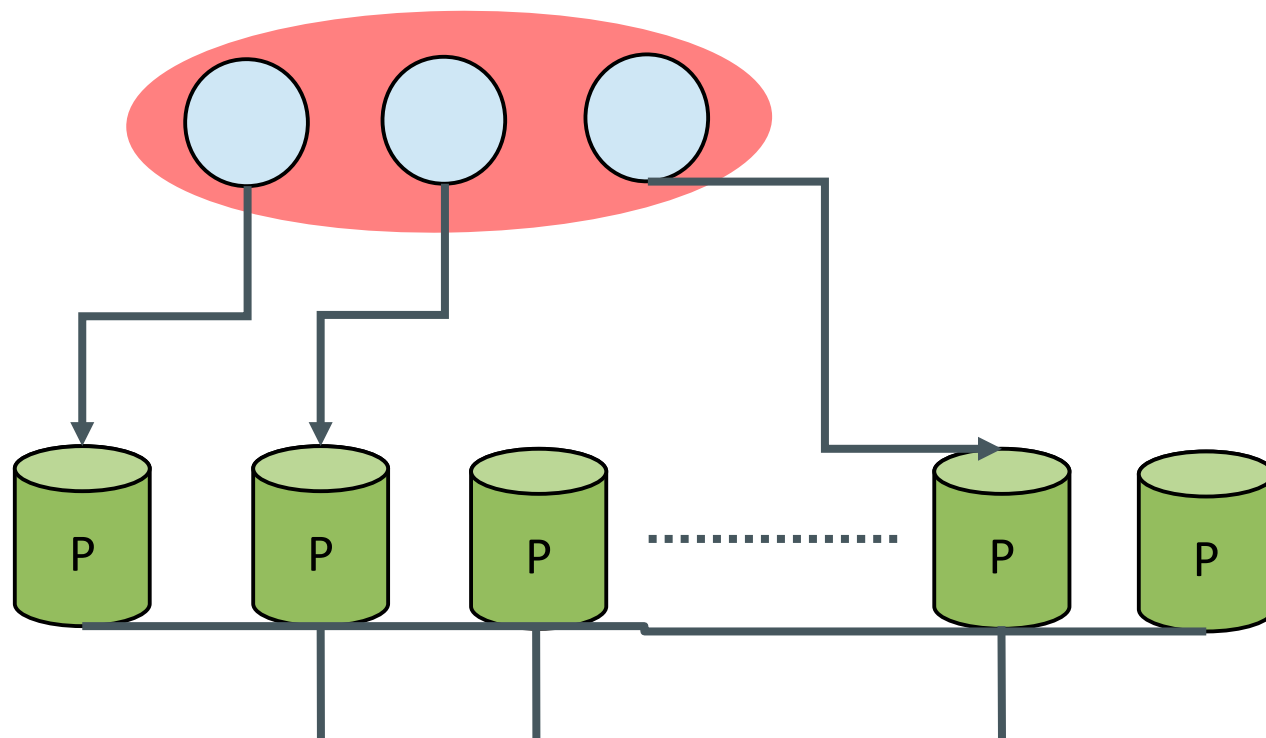


2 Примеры использования

Примеры использования

- **Elastic Replication**

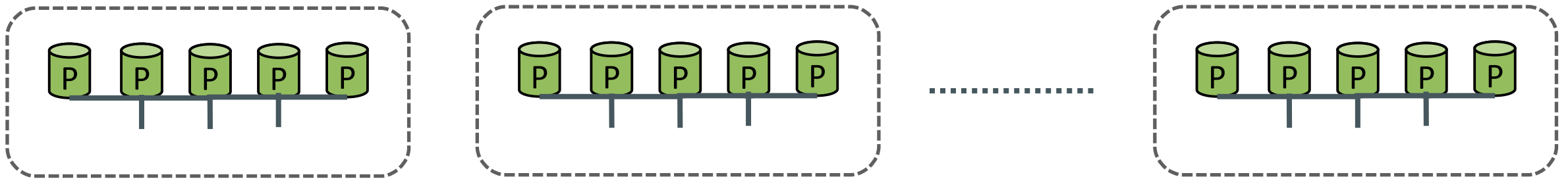
- Окружение с часто изменяющимся количеством серверов репликаций и с необходимостью максимально упростить процесс этих изменений.



Примеры использования

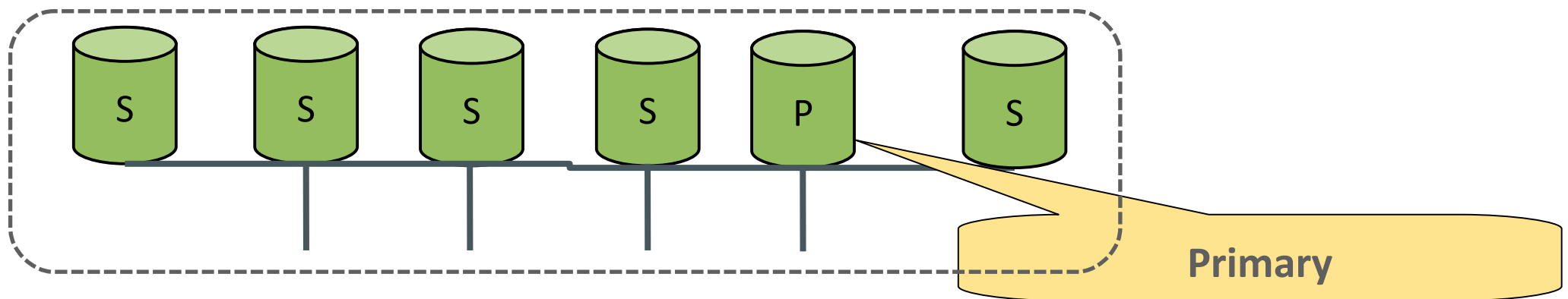
- **Highly Available Shards**

- Шардинг является популярным способом масштабирования нагрузки по изменению данных. MySQL Group Replication может быть использован для создания таких отказоустойчивых групп.



Примеры использования

- Альтернатива стандартной схеме репликации Master-Slave
- **Single-primary режим предоставляет дополнительные возможности:**
 - Автоматическое определение ролей PRIMARY/SECONDARY
 - Автоматический выбор нового Primary сервера в случае отказа предыдущего
 - Автоматическая настройка соответствующих режимов RO/RW



Режимы работы

3

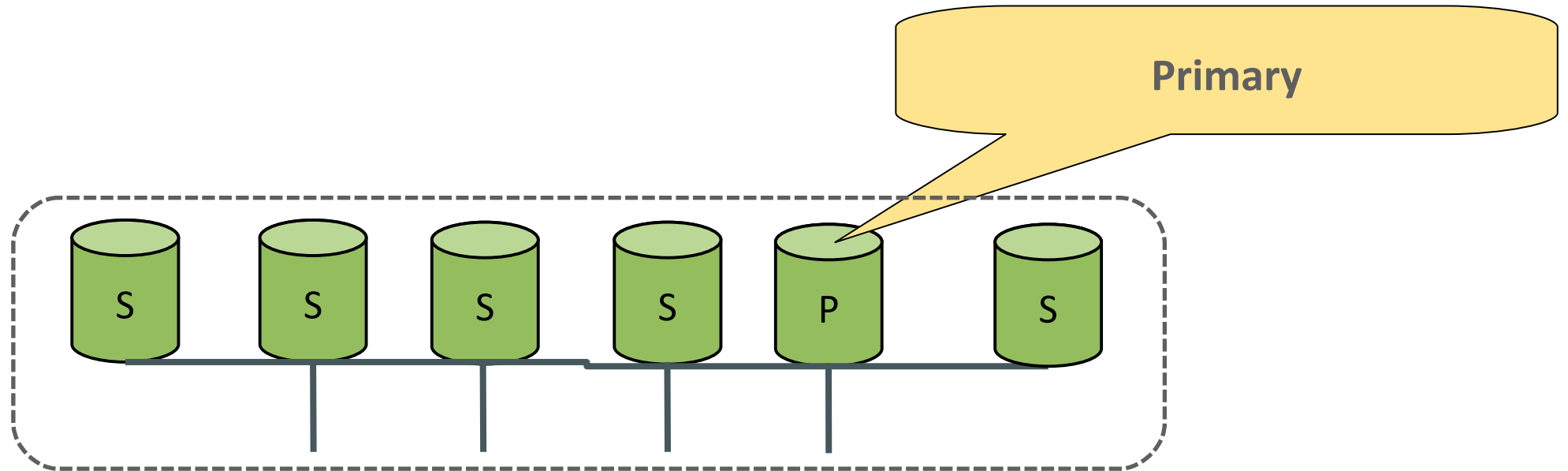
3.1 Single-primary

Single-primary

- Режим в котором один сервер принимает запросы на изменения данных (PRIMARY), а остальные доступны для чтения, всегда содержат актуальные данные и при необходимости могут быть использованы для отказоустойчивости (SECONDARIES).
 - Группа сама выбирает Primary сервер с помощью специального механизма выборов.
- Single-primary является режимом по умолчанию.
 - Ближе к стандартной асинхронной репликации.
 - Не имеет некоторых ограничений multi-primary режима.

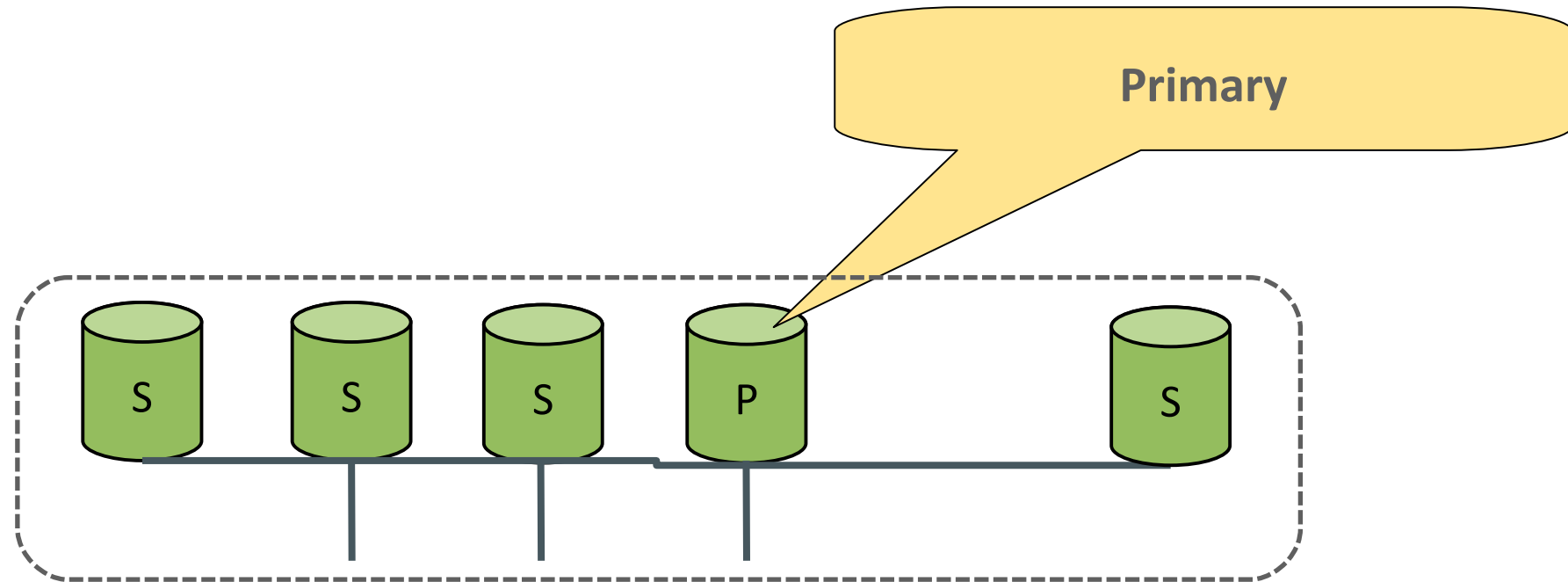
Single-primary

- Автоматический выбор нового Primary сервера.
- Secondary сервера автоматически переводятся в режим read-only.



Single-primary

- Автоматический выбор нового Primary сервера.



Single-primary

- UUID текущего Primary сервера можно определить следующим запросом.

```
mysql> SELECT * FROM performance_schema.global_status WHERE  
        VARIABLE_NAME='group_replication_primary_member';  
VARIABLE_NAME          VARIABLE_VALUE  
group_replication_primary_member  dcd3b36b-79c5-11e6-97b8-00212844d44e
```

Режимы работы

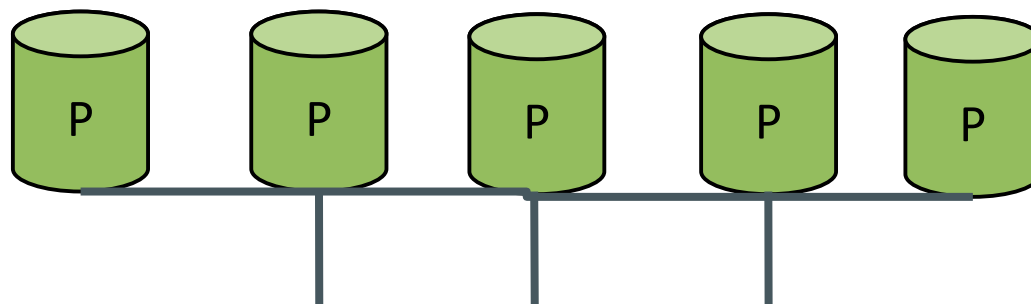
3

3.1 Single-primary

3.2 Multi-primary

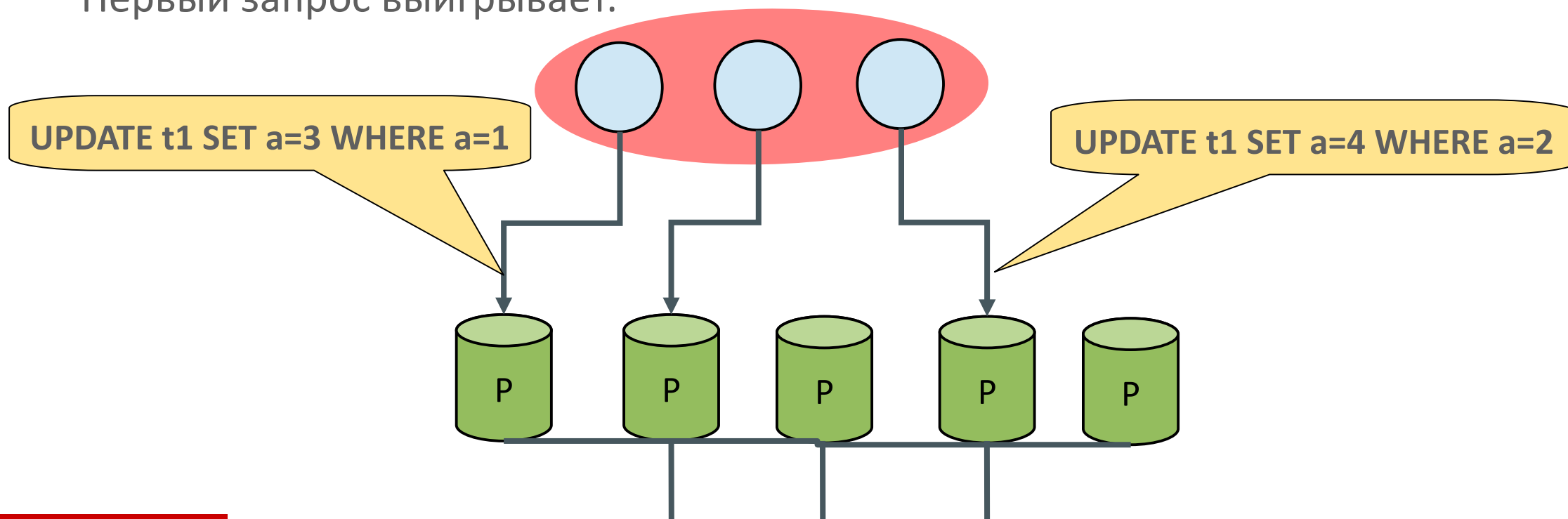
Multi-primary update everywhere!

- Режим в котором запись возможна на любом сервере
 - Включается изменением значения переменной `--group_replication_single_primary_mode=OFF`



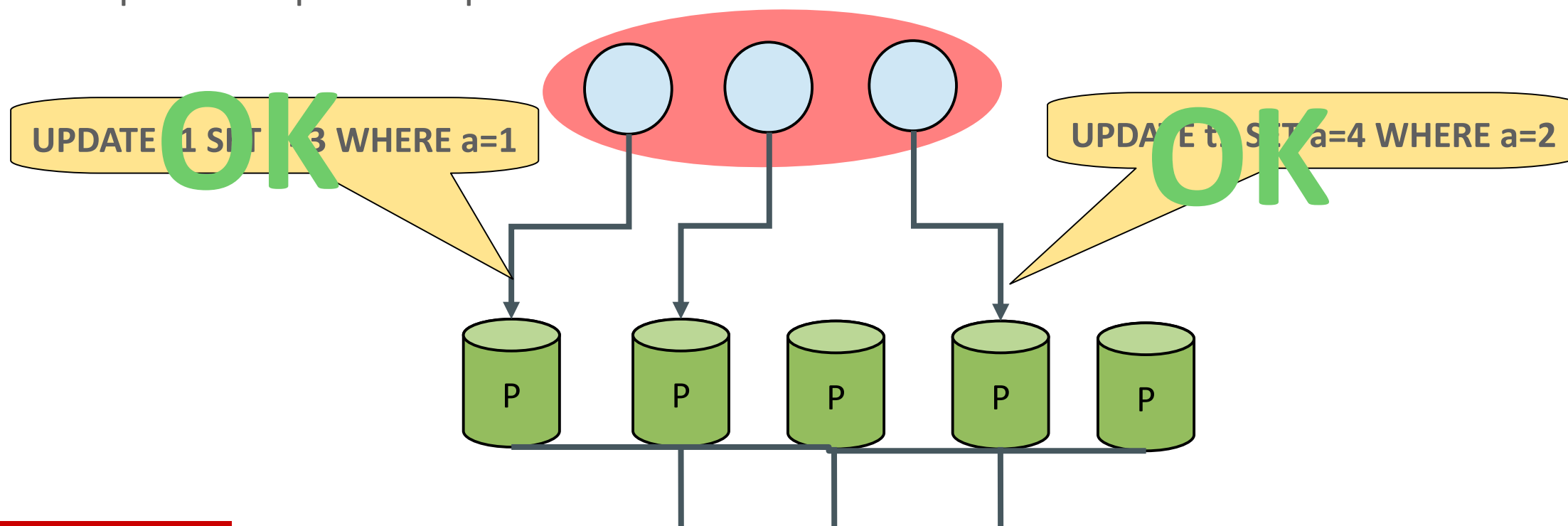
Multi-primary update everywhere!

- Любые транзакции могут изменять любые данные.
- Конфликты будут обнаружены и разрешены.
 - Первый запрос выигрывает.



Multi-primary update everywhere!

- Любые транзакции могут изменять любые данные.
- Конфликты будут обнаружены и разрешены.
 - Первый запрос выигрывает.

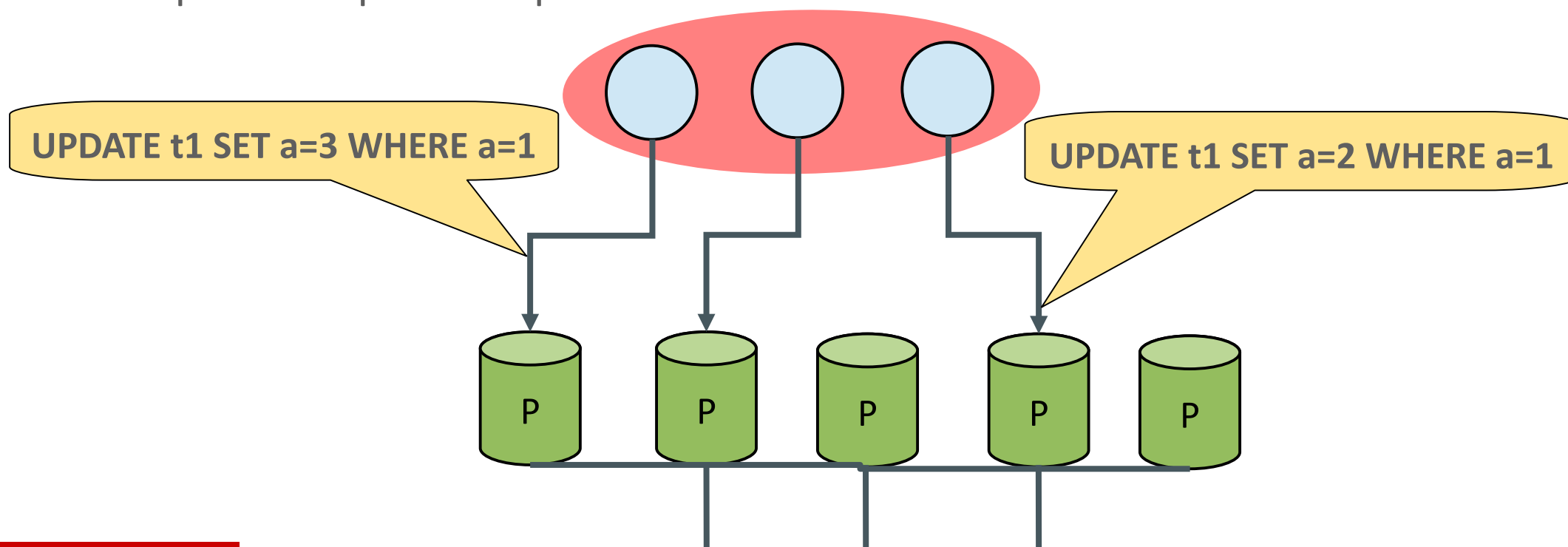


Multi-primary update everywhere!

Любые транзакции могут изменять любые данные.

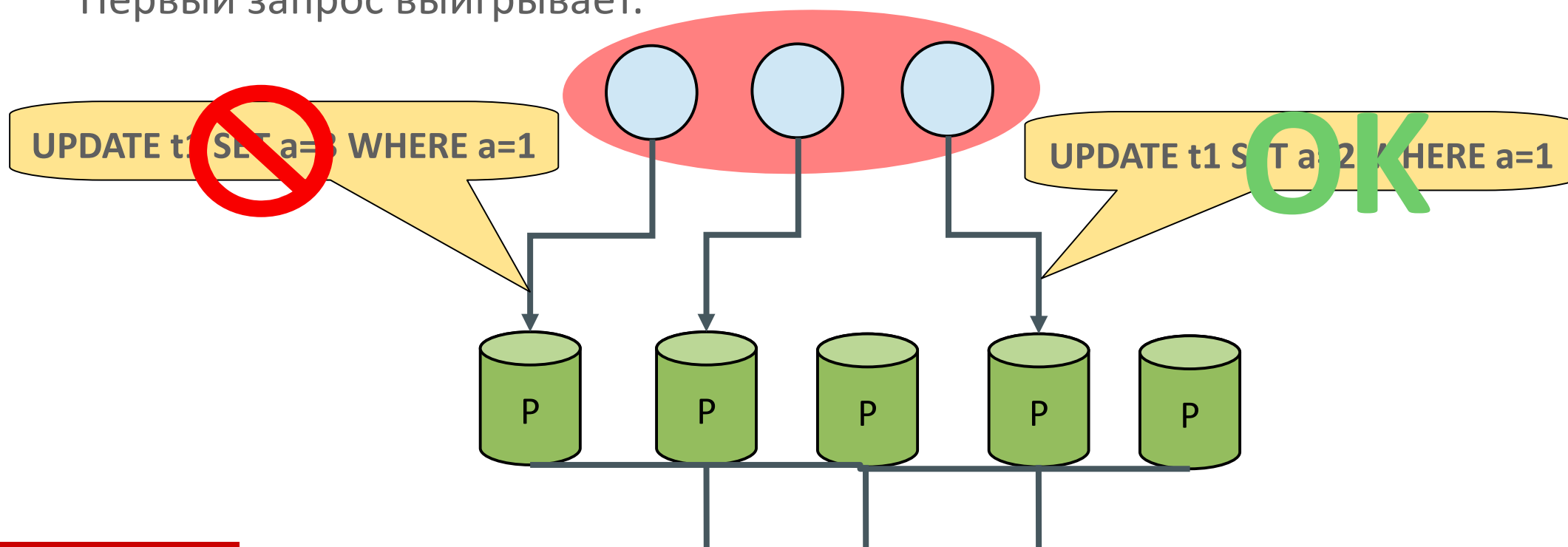
Конфликты будут обнаружены и разрешены.

- Первый запрос выигрывает..



Multi-primary update everywhere!

- Любые транзакции могут изменять любые данные.
- Конфликты будут обнаружены и разрешены.
 - Первый запрос выигрывает.



Features

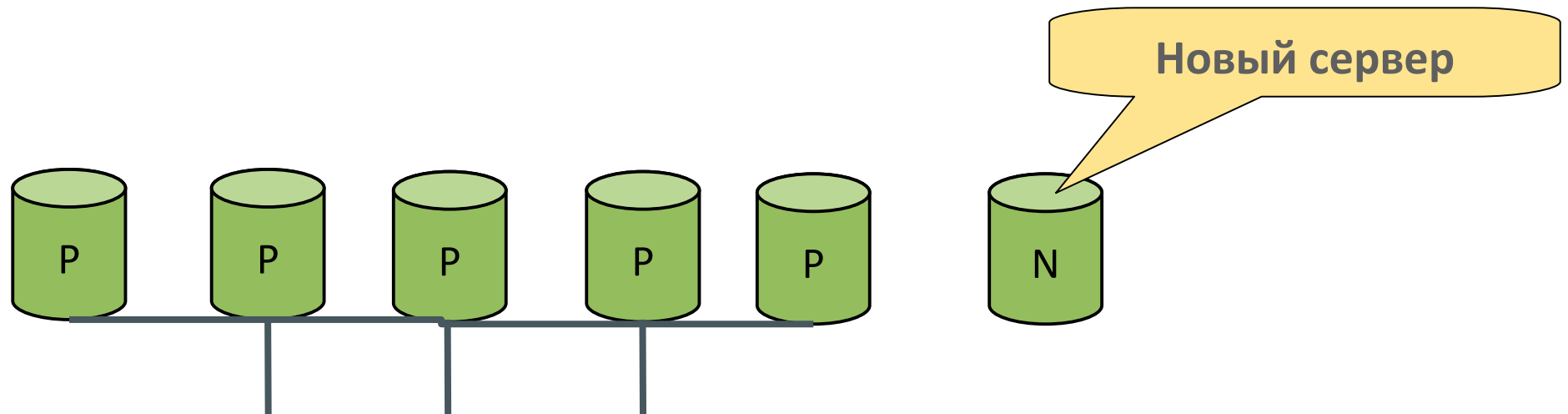
4

4.1

Автоматическое восстановление серверов

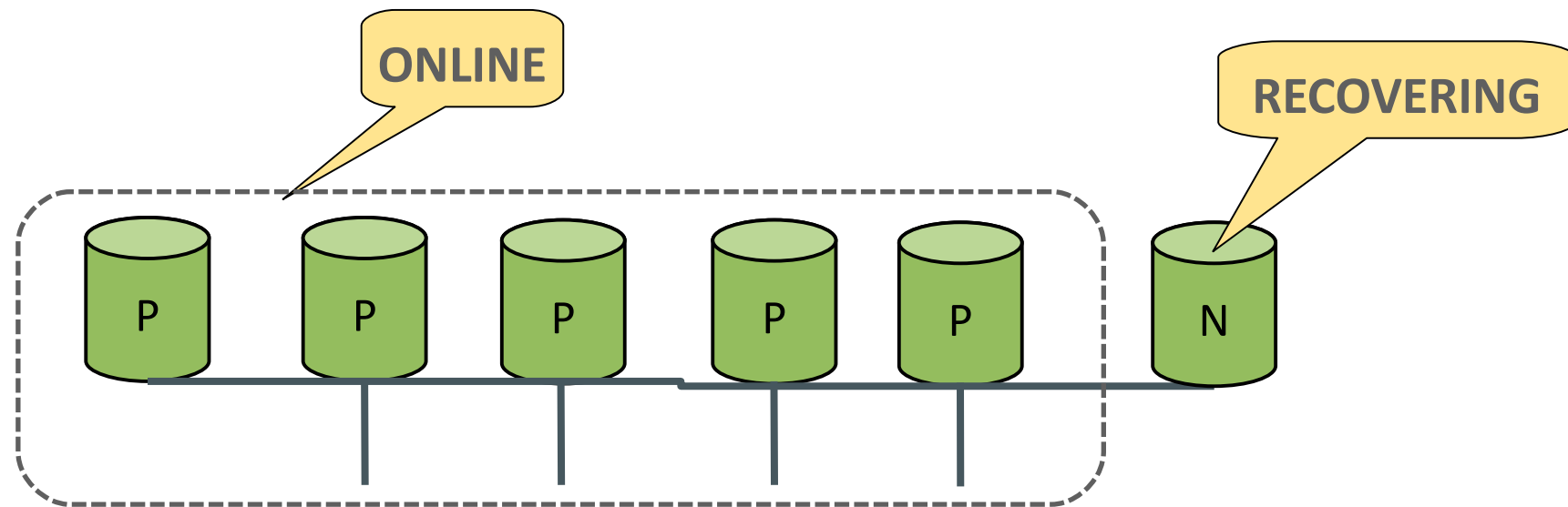
Автоматическое восстановление серверов

- Сервер, который присоединяется к группе, автоматически синхронизирует данные.



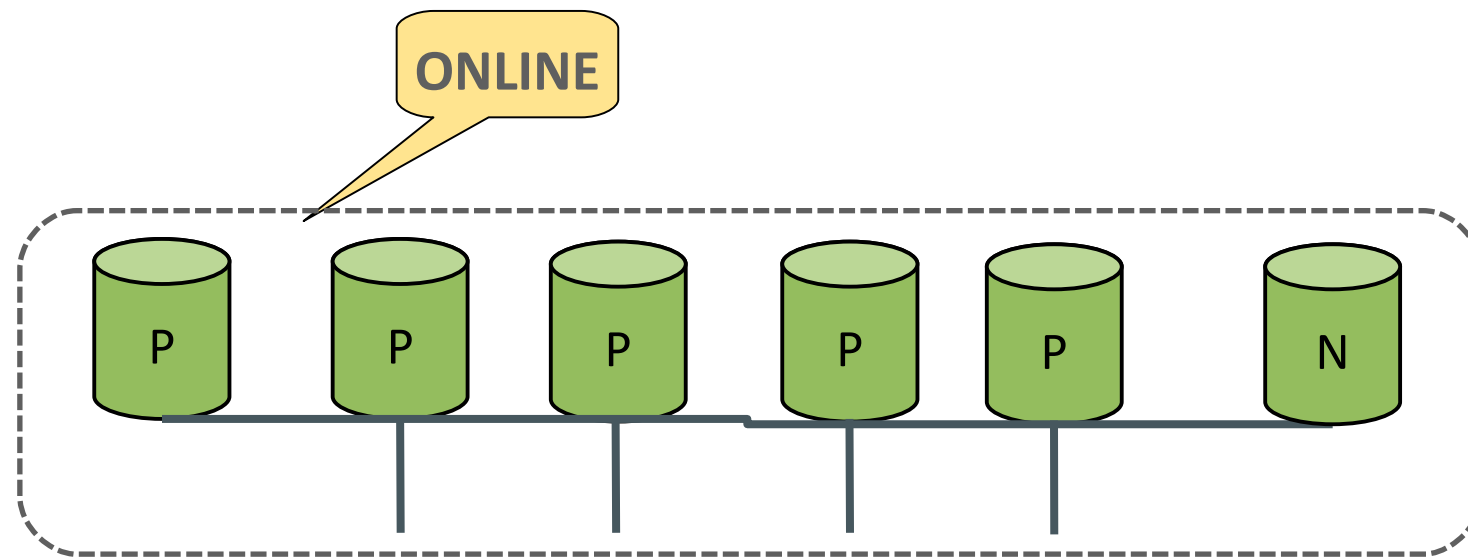
Автоматическое восстановление серверов

Сервер, который присоединяется к группе, автоматически синхронизирует данные.



Автоматическое восстановление серверов

Сервер, который присоединяется к группе, автоматически синхронизирует данные.

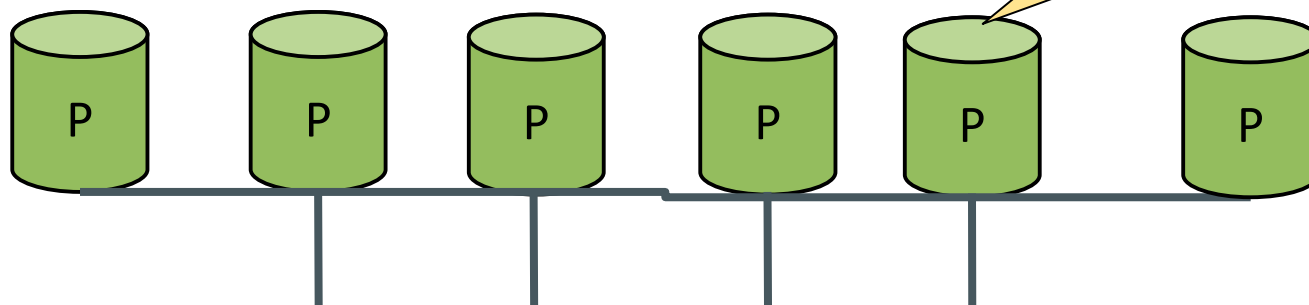


Автоматическое восстановление серверов

- Если сервер покидает группу, то все остальные серверы в группе будут об этом проинформированы.

Каждая конфигурация серверов определяется `view_id`

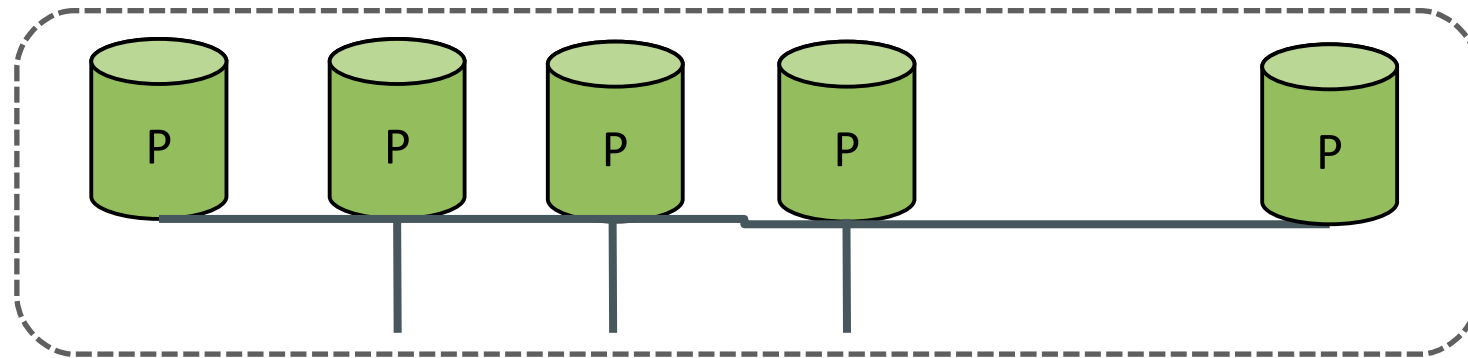
Этот сервер отключается на обслуживание



`view_id: 4`

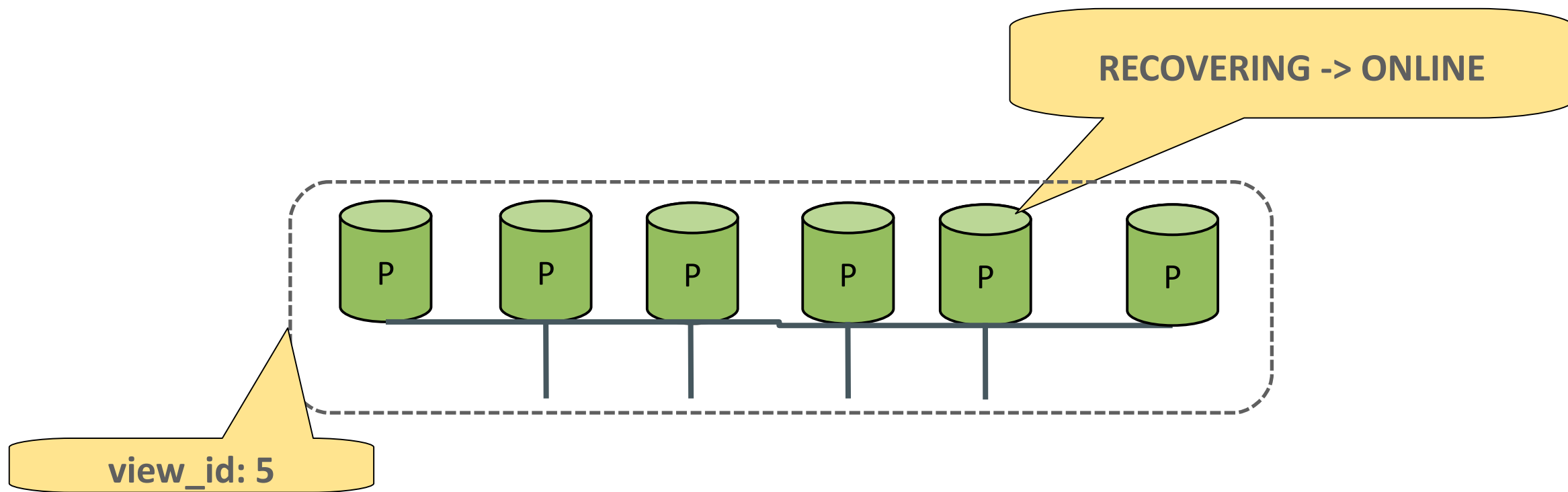
Автоматическое восстановление серверов

- Если сервер покидает группу, то остальные сервера будут об этом автоматически проинформированы.



Автоматическое восстановление серверов

- Когда сервер снова возвращается в группу он автоматически синхронизируется с остальной группой.



Features

4

4.1 Автоматическое восстановление серверов

4.2 Интеграция с MySQL

Интеграция с MySQL

- MySQL Plugin
 - Стандартный MySQL плагин.
- MySQL InnoDB
 - Работа с данными такая же как и всегда для InnoDB.
 - Оптимизация InnoDB для лучшей поддержки Group Replication.
- MySQL Performance Schema
 - Мониторинг Group Replication осуществляется с помощью таблиц в Performance Schema.

Интеграция с MySQL

- Group Replication не использует внешних компонентов.
- Для использования Group Replication требуются те же навыки, что и для работы с самим MySQL Server.

Features

4

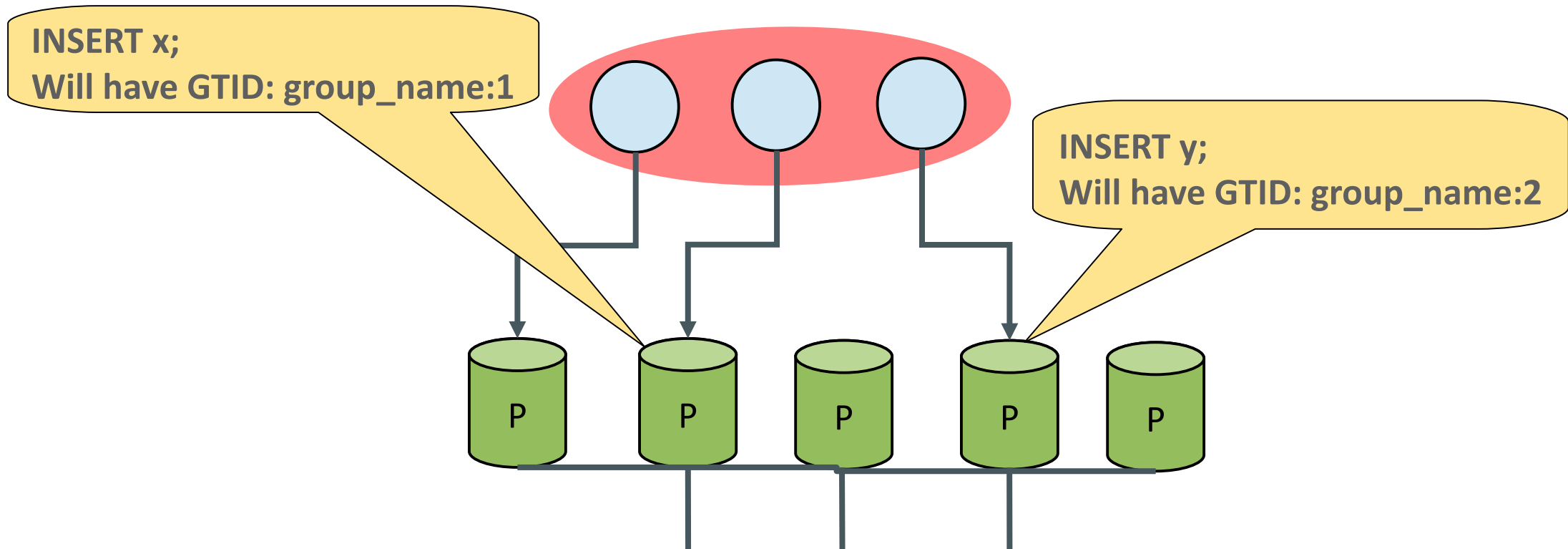
4.1 Автоматическое восстановление серверов

4.2 Интеграция с MySQL

4.3 Полная поддержка GTID

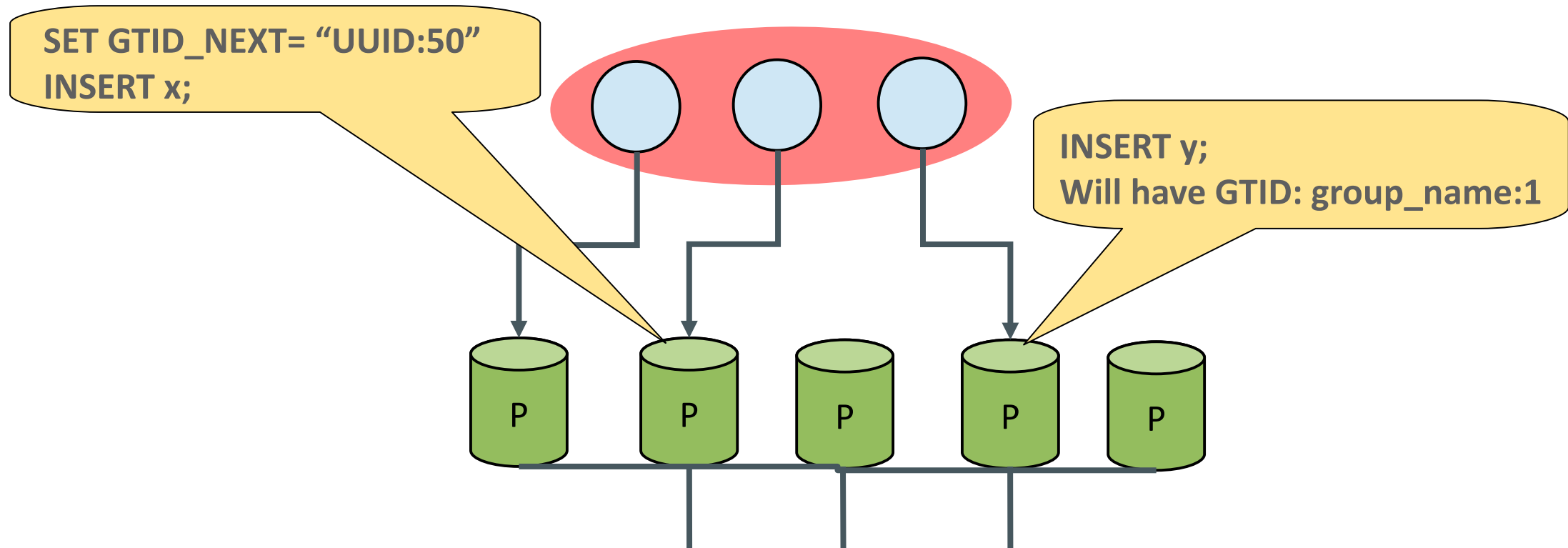
Полная поддержка GTID

- У всех транзакций в группе одинаковый UUID соответствующий ID группы.



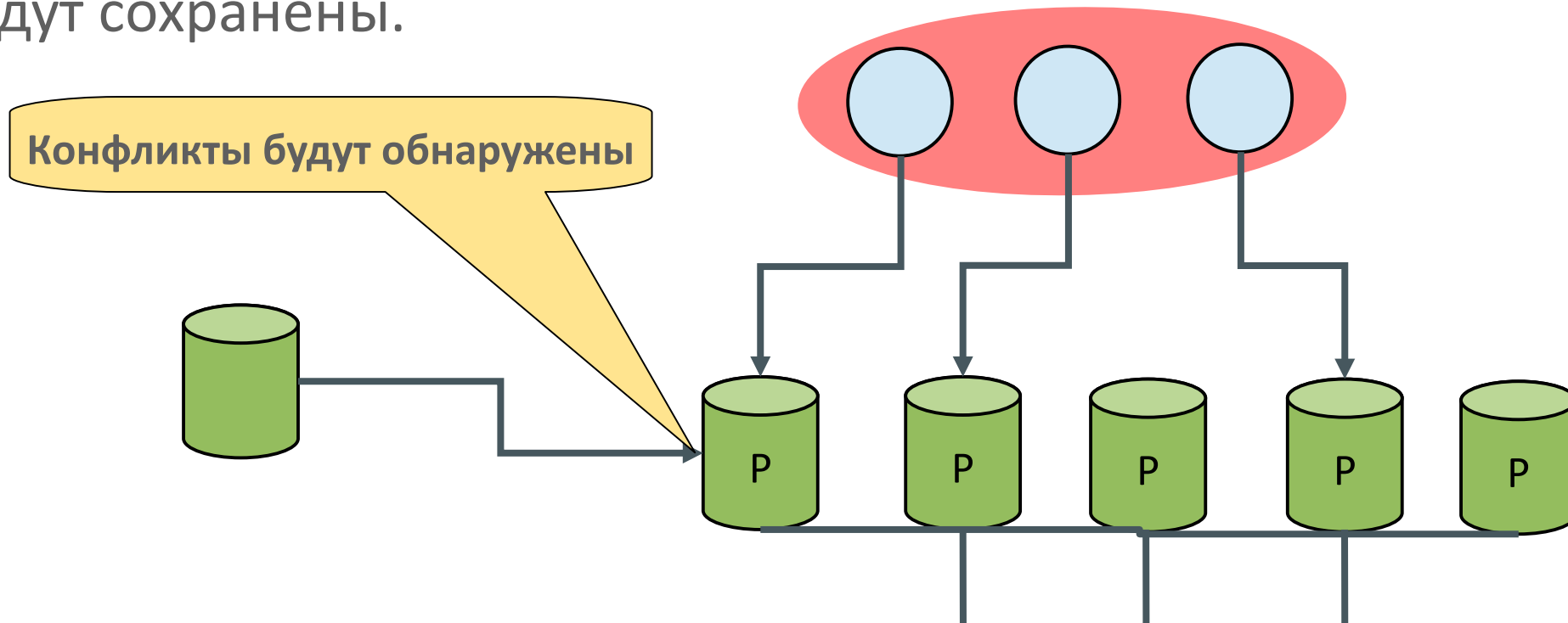
Полная поддержка GTID

- Возможна ручная настройка идентификаторов пользователем.



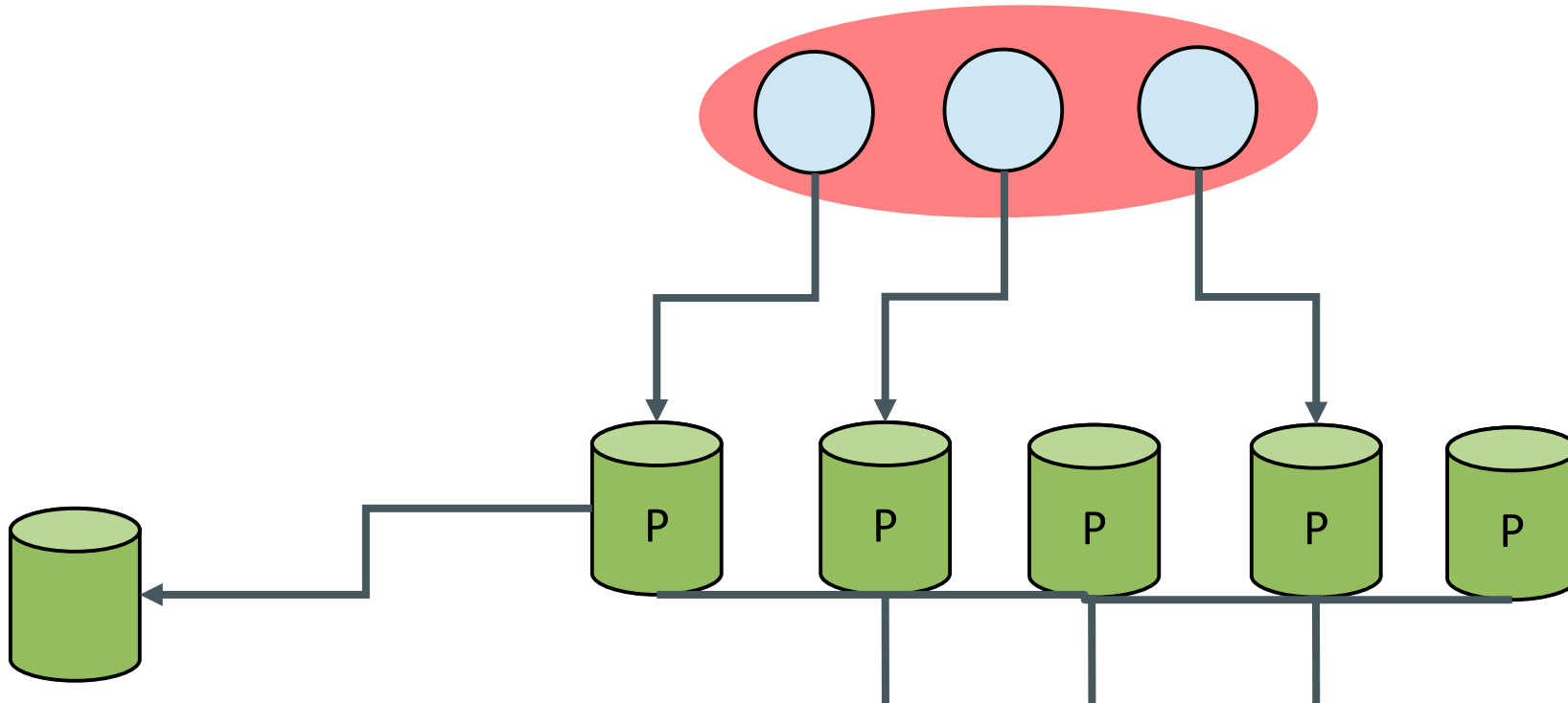
Полная поддержка GTID

- Есть возможность репликации от внешнего сервера к группе, GTID будут сохранены.



Полная поддержка GTID

- Есть возможность репликации из группы к внешнему серверу.



Features

4

4.1 Автоматическое восстановление серверов

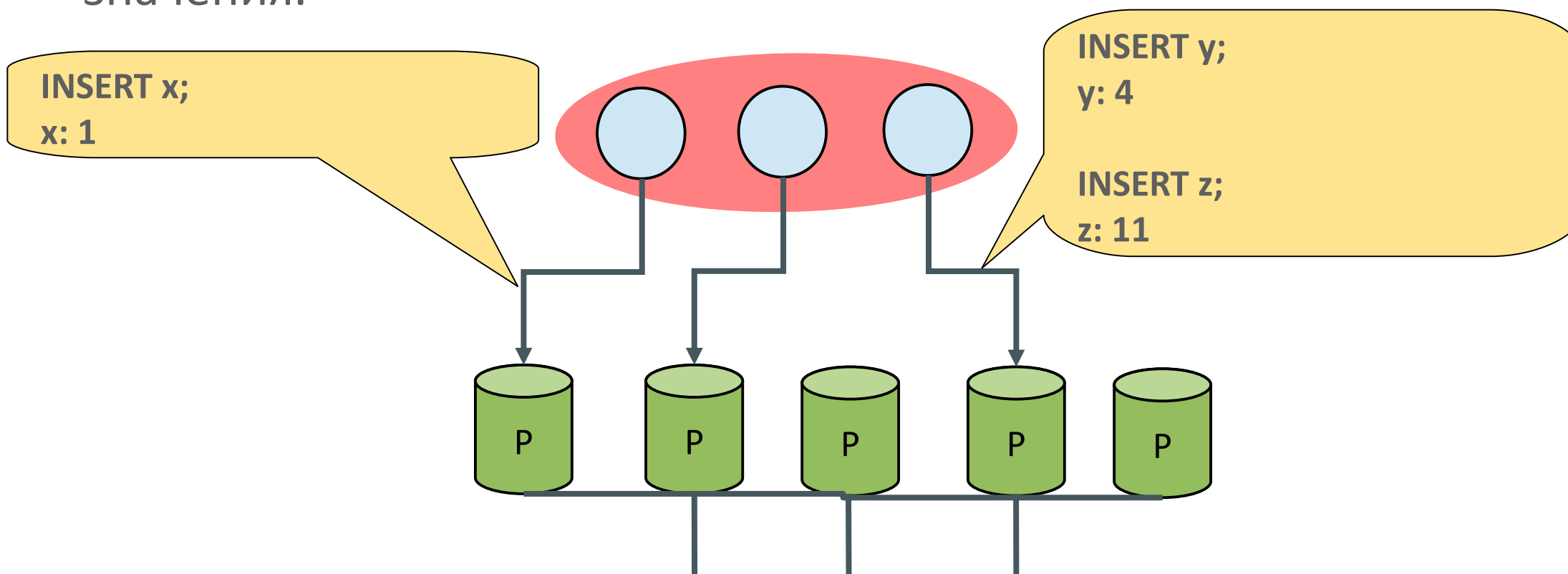
4.2 Интеграция с MySQL

4.3 Полная поддержка GTID

4.4 Настройка auto-increment

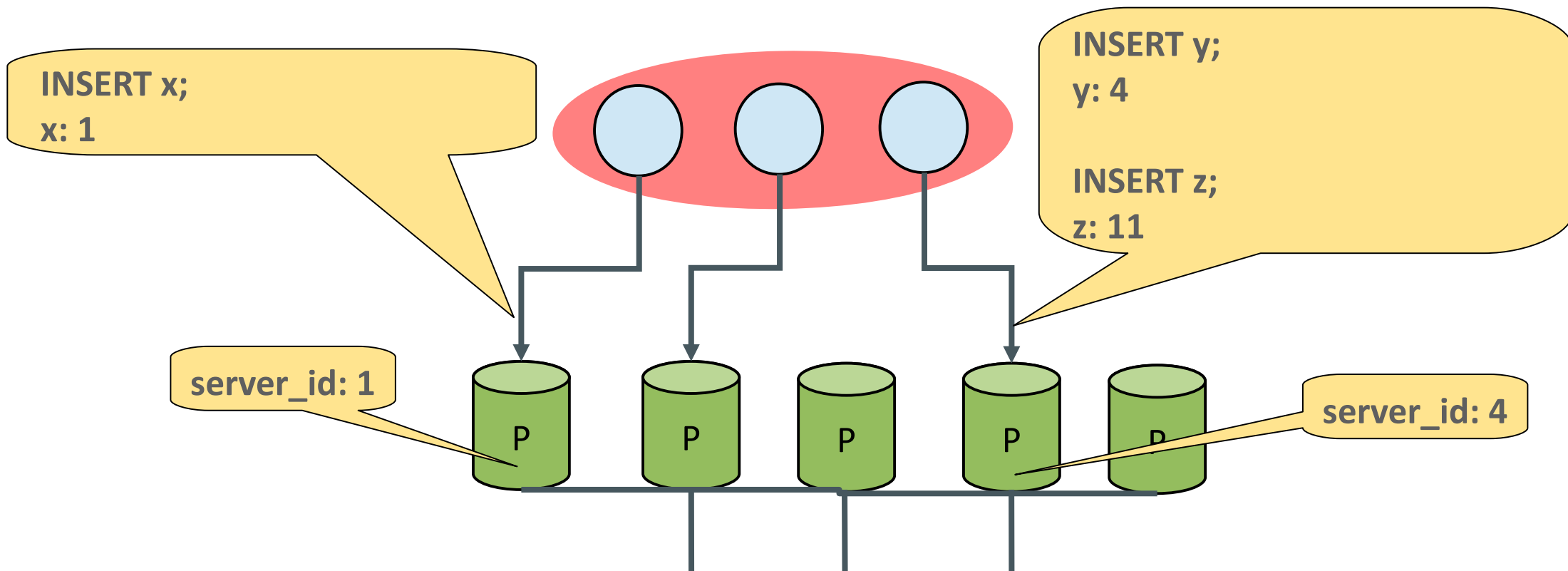
Настройка auto-increment

- Группа настроена не генерировать одинаковые auto-increment значения.



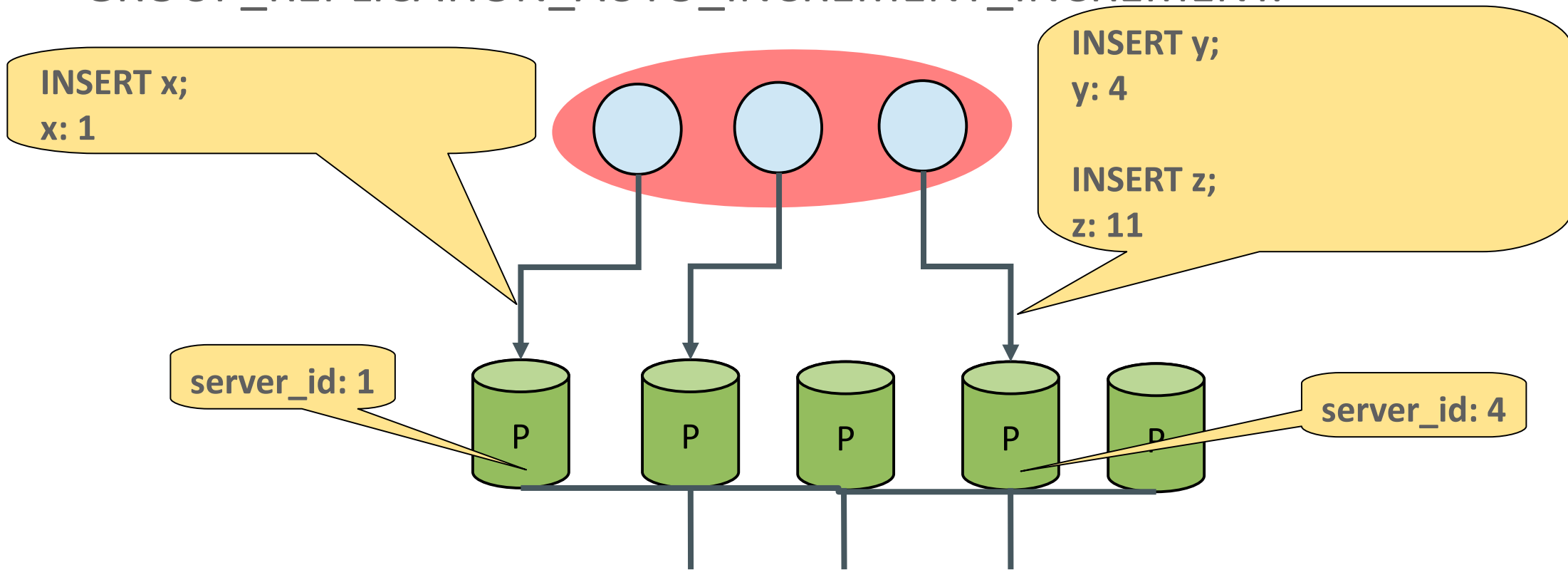
Настройка auto-increment

- По умолчанию сдвиг определяется по `server_id` и увеличивается на 7.



Auto-increment configuration/handling

- Пользователь может изменить сдвиг с помощью опции GROUP_REPLICATION_AUTO_INCREMENT_INCREMENT.



Features

4

4.1 Автоматическое восстановление серверов

4.2 Интеграция с MySQL

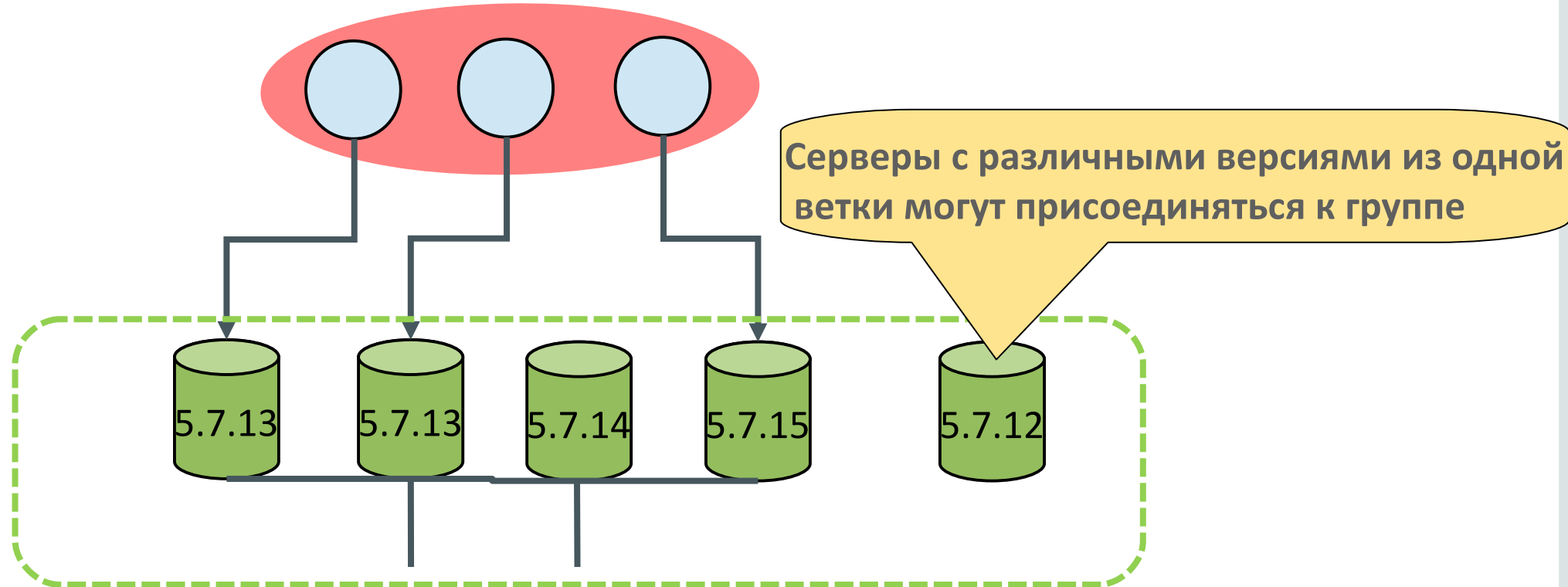
4.3 Полная поддержка GTID

4.4 Настройка auto-increment

4.5 Поддержка различных версий

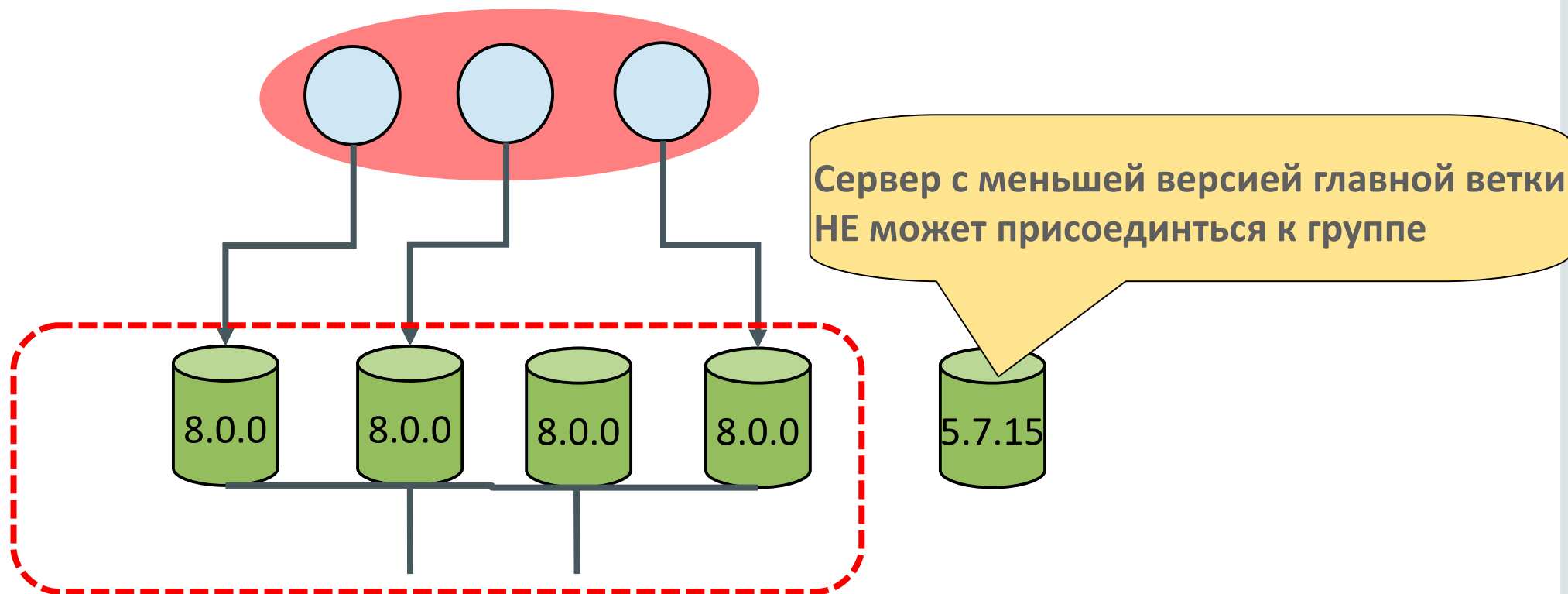
Поддержка различных версий

- При присоединении нового сервера, его версия важна для совместимости с группой.



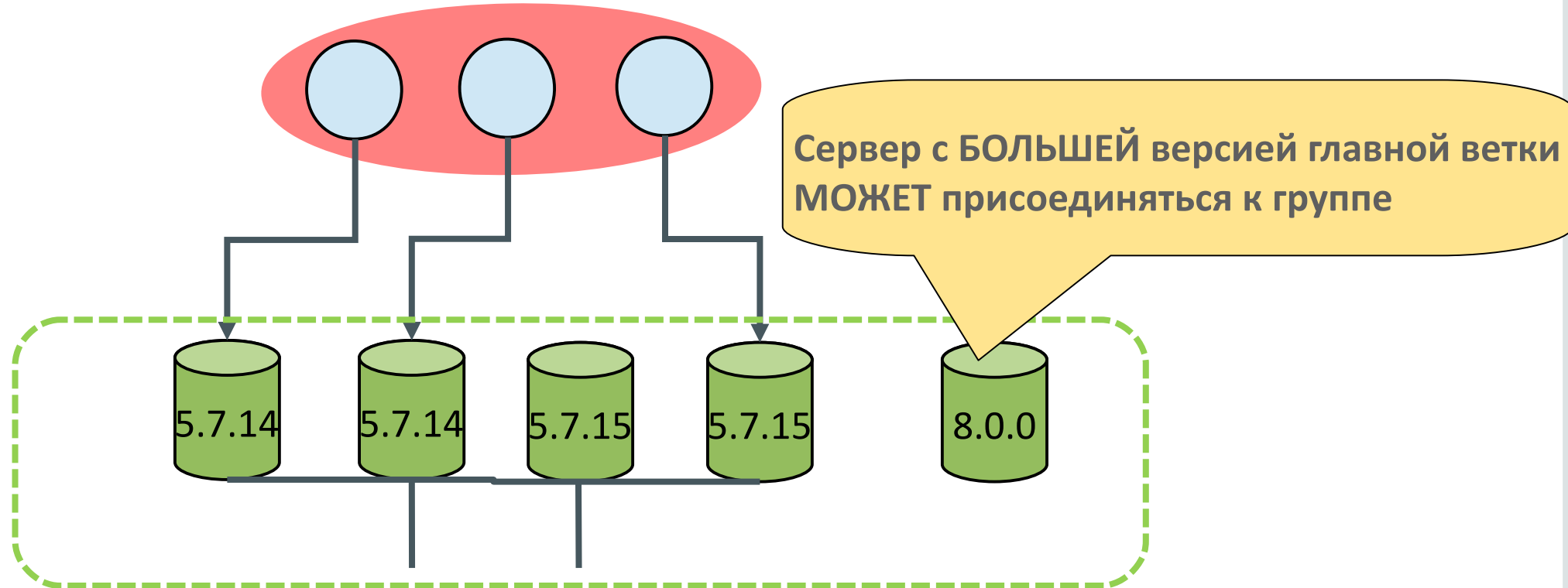
Поддержка различных версий

При присоединении нового сервера, его версия важна для совместимости с группой.



Поддержка различных версий

При присоединении нового сервера, его версия важна для совместимости с группой.



Features

4

4.1 Automatic distributed server recovery

4.2 Интеграция с MySQL

4.3 Полная поддержка GTID

4.4 Настройка auto-increment

4.5 Поддержка различных версий

4.6 Встроенный communication engine

Встроенный Communication Engine

- Плагин основан на семействе протоколов Paxos.
 - Compression, multi-platform, dynamic membership, distributed agreement, quorum based message passing, SSL, IP whitelisting.
- Не требуются никакие third-party приложения.
- Не требуются сети с поддержкой multicast.
 - MySQL Group Replication может работать в облаке, где multicast не всегда поддерживается.

Features

4

4.1 Automatic distributed server recovery

4.2 Интеграция с MySQL

4.3 Полная поддержка GTID

4.4 Настройка auto-increment

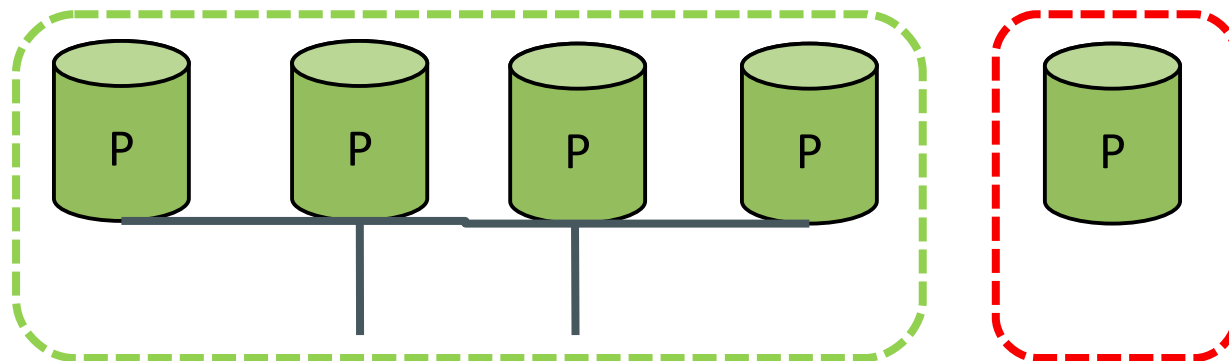
4.5 Поддержка различных версий

4.6 Встроенный communication engine

4.7 Режим read-only

Режим Read-only

- Когда новый сервер присоединяется к группе, на время процесса синхронизации данных автоматически устанавливается режим read-only.
- В случае ошибки на одном из серверов, он автоматически меняет статус на ERROR и переходит в режим read-only, чтобы приложение по ошибке не могло изменить данные и нарушить целостность данных.



Features

4

4.8 Поддержка безопасных подключений

Поддержка безопасных подключений

- Group Replication поддерживает безопасные подключения:
 - Соединения для восстановления
 - Соединения между различными серверами в группе
 - Соединения между клиентами и группой
- IP Whitelisting
 - Список IP адресов, с которых разрешено присоединяться к группе.
 - По умолчанию стоит значение AUTOMATIC, которое позволяет присоединяться из private подсети соответствующей текущему хосту.

Features

4

4.8 Поддержка безопасных подключений

4.9 Поддержка Parallel applier

Поддержка Parallel applier

- Значительно увеличивает производительность репликации.
- Настройка аналогична асинхронной репликации.

```
--slave_parallel_workers=NUMBER  
--slave_parallel_type=logical_clock  
--slave_preserve_commit_order=ON
```

Поддержка Parallel applier

- Определение Transaction Dependencies на основе Write Set
 - Изначально используется в Group Replication
 - Ускоряет процесс восстановления сервера

```
master> SET @@GLOBAL.binlog_transaction_dependency_tracking=WRITESET;  
Query OK, 0 rows affected (0,00 sec)
```

```
master> SET @@GLOBAL.binlog_transaction_dependency_tracking=WRITESET_SESSION;  
Query OK, 0 rows affected (0,00 sec)
```

```
master> SET @@GLOBAL.binlog_transaction_dependency_tracking=COMMIT_ORDER; -- default  
Query OK, 0 rows affected (0,00 sec)
```

Features

4

4.8 Поддержка безопасных подключений

4.9 Поддержка Parallel applier

4.10 Поддержка Transaction SAVEPOINT

Поддержка Transaction SAVEPOINT

```
mysql> BEGIN;  
Query OK, 0 rows affected (0,00 sec)
```

```
mysql> INSERT INTO t1 VALUES(1);  
Query OK, 1 row affected (0,00 sec)
```

```
mysql> SAVEPOINT s1;  
Query OK, 0 rows affected (0,00 sec)
```

```
mysql> INSERT INTO t1 VALUES(2);  
Query OK, 1 row affected (0,00 sec)
```

```
mysql> ROLLBACK TO s1;  
Query OK, 0 rows affected (0,00 sec)
```

```
mysql> COMMIT;  
Query OK, 0 rows affected (0,00 sec)
```

Features

4

4.8 Поддержка безопасных подключений

4.9 Поддержка Parallel applier

4.10 Поддержка Transaction SAVEPOINT

4.11 Требования

Требования (by design)

- Требуется InnoDB storage engine
- Нужен primary key в каждой таблице
- Требуются GTIDs
- Требуется бинарный лог
- Требуется row format binlog-a
- Optimistic execution: транзакция может быть отменена в момент COMMIT-а из-за конфликта с другой транзакцией на других серверах группы
- До 9 серверов в группе

Не поддерживается

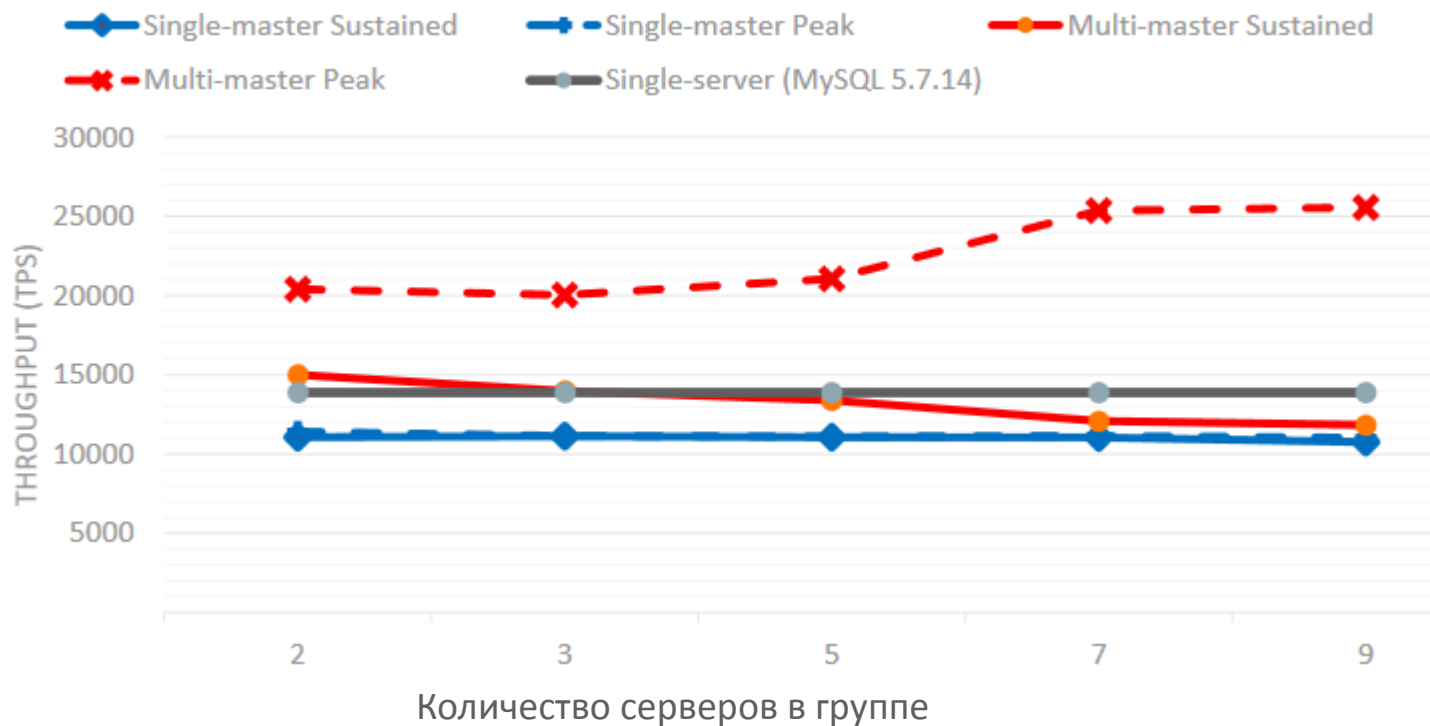
- Serializable (для multi-primary)
- Cascading Foreign Keys (для multi-primary)
- Binary log events checksum

Warnings

- Одновременные DDL (для multi-primary)
- SELECT *** FOR UPDATE не использует group locking (для multi-primary)

5 Производительность

Производительность



Peak Throughput (i.e., no flow control)

Количество записе, которые приложение может сделать (в секунду).

Sustained Throughput (i.e., flow control)

Количество транзакций, которое приложение может записать в группу без отставания серверов (в секунду).

Servers

9 Dual Xeon E5-2660-v3
Enterprise SSD Storage
10Gbps Ethernet Network

Client

1 Dual Xeon E5-2699-v3
10Gbps Ethernet Network
Sysbench 0.5 RW workload

More on this subject on the series of replication performance blogs at:
<http://mysqlhighavailability.com/category/performance/>

Производительность

- **Усреднённая пропускная способность:**
 - Производительность в Multi-primary режиме плавно снижается с ростом количества серверов в группе.
 - Производительность в Single-primary режиме падает минимально с ростом количества серверов в группе.
- **Пиковая пропускная способность:**
 - Производительность в Multi-primary режиме может увеличиваться до 1.8 раз в сравнении с одиночным сервером.
 - Нагрузка по чтению данных распределена между серверами.
 - В single-primary режиме отставание на secondary серверах отсутствует.

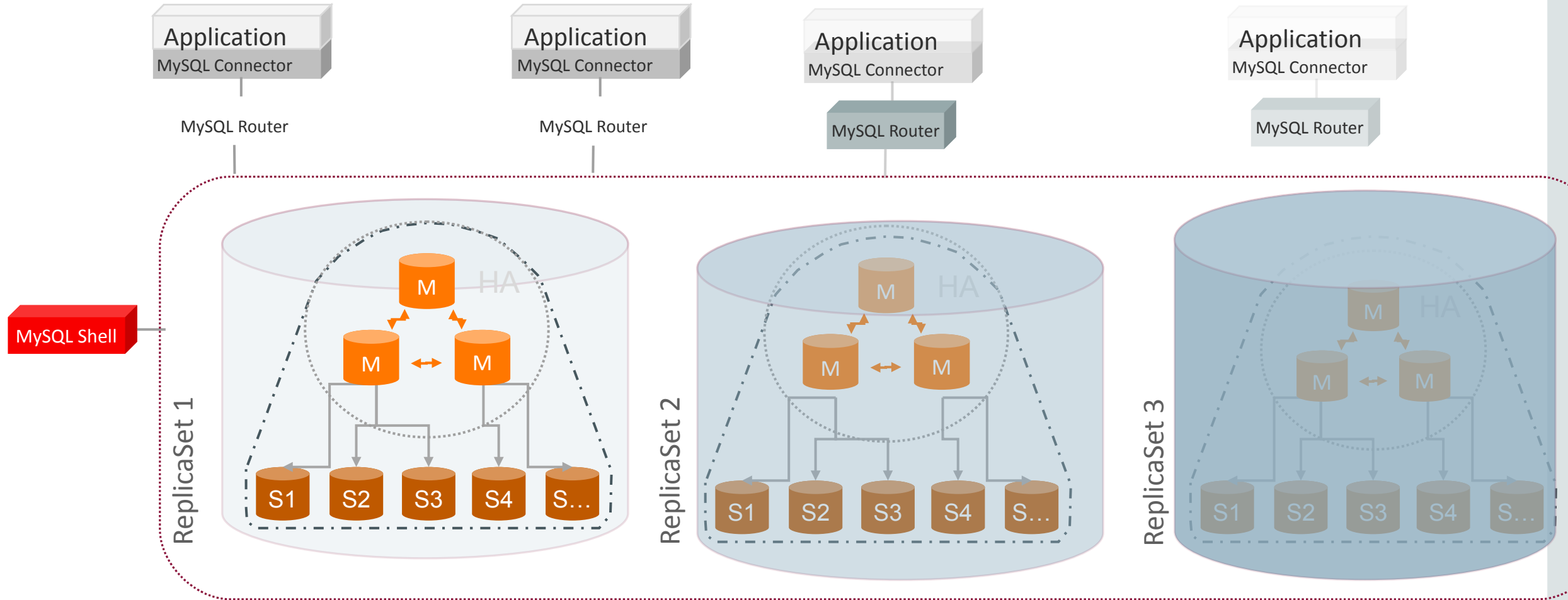
Заключение



Заключение

- **Cloud Friendly**
 - Подходит для систем с часто изменяющимся составом серверов.
- **Интегрирован**
 - С ядром MySQL Server.
 - С GTID, row based replication, таблицами performance schema.
- **Автономный и упрощенный в использовании**
 - Самовостановление: не требуется вмешательство админа для переключения.
 - Предоставляет отказоустойчивость, позволяет построить “update anywhere” систему.
- **GA версии** плагина доступны в MySQL 5.7.17+ и в 8.0.1+

MySQL InnoDB Cluster: **The End Goal**



Дополнительные ссылки

- Доступно для скачивания:
 - <http://www.mysql.com/downloads/>
- Документация:
 - <http://dev.mysql.com/doc/refman/5.7/en/group-replication.html>
 - <http://dev.mysql.com/doc/refman/8.0/en/group-replication.html>
- Блоги разработчиков:
 - <http://mysqlhighavailability.com>

ORACLE®