

srv-postgre.olvr: standby только для DR — ограничения collation и recovery

Задача: [OLV-4416](#). Единственный carrier этого runbook — `runbooks:srv-postgre-olvr`; он заменяет прежний указатель `srv-postgre.olvr.md` из `sysmanage`.

Статус: DR-only. Standby не квалифицирован как горячая замена primary. Collation mismatch не устранён; готовность failover и приемлемый численный RTO не доказаны.

Этот документ описывает ограничения и условную будущую recovery-операцию. Его публикация **не разрешает** promotion, REINDEX, REFRESH, изменение БД, доступов или конфигурации. Реальная операция требует назначения оператора, проверки target и отдельного разрешения владельца на production writes.

Targets и сохранённые факты

- Standby: `srv-postgre.olvr`, 192.168.176.9:5432, socket `/var/tmp/pgsql-standby`; PostgreSQL 17.9, FreeBSD 14.3.
- Primary: `srv-prod-postgre.olvr`; установленный hostname `pgsql-ng`, Debian 11, glibc 2.31-13+deb11u6.
- Сохранённый mismatch: `datcollversion=2.31`, версия collation на standby — 43.0. WARNING при подключении сообщает об этом риске.
- На `srv-postgre.olvr` есть также тестовый инстанс: вызов `psql` без явного endpoint может обратиться не к standby. Для standby использовать только подтверждённый endpoint; имя host само по себе не подтверждает инстанс.

Факты — датированные результаты DESIGN Yqlv/gvQSwN и принятых измерений NGw4/9jyCrC от 02.10.2026, а не новая проверка состояния сервера при чтении этого документа.

Ограничение чтения

Не использовать standby для сверок, аналитики и любых выборок, где важна корректность результата по текстовым ключам или текстовому порядку. Индексы построены с правилами сортировки primary; библиотека standby может иначе упорядочивать те же ключи. Возможны неверные результаты поиска и порядка.

Вместо standby использовать primary через существующий разрешённый read-only доступ (например, роль `monitoring`, если доступ подтверждён для нужной базы) либо разрешённую выгрузку с primary. Этот runbook новых прав доступа не выдаёт.

Технический запрет через `pg_hba` и автоматическое отключение мониторинга не вводятся. Подтверждённый потребитель standby — локальный `postgres_exporter`, роль `prometheus`, соединение с явным standby socket. Его наличие не означает, что все периодические читатели установлены.

Операционный получатель уведомления — РП **OLVR**: ограничения приняты, мониторинг не отключать. Внешний владелец exporter отдельно не установлен; этот gap передан OLVR без нового inventory.

Отсутствие WARNING после REFRESH не доказывает корректности индексов или размещения строк. REFRESH меняет метаданные версии, но зависимые объекты не исправляет.

Принятая область затронутой collation

Определение одинаково для индексов и range-ключей:

- `pg_collation.collprovider = 'c'` (libc), locale не C/POSIX;
- `collprovider = 'i'` (ICU);
- `collprovider = 'd'` (default базы), если `pg_database.dattlocprovider = 'i'` либо `dattlocprovider = 'c'` с `datcollate` не C/POSIX.

Не входят: provider b (builtin), C/POSIX и не-collatable ключи.

ICU включена намеренно: равенство библиотек ICU на primary и standby не доказано. Это сохранённая fail-closed граница принятого DESIGN, а не разрешение переходить на ICU или поручение проверять её версии. Новый алгоритм классификации здесь не вводится.

Измеренный вход и предел RTO

В 30 измеренных базах, включая template1:

- **297** btree-индексов с затронутой collation;
- суммарный `pg_relation_size` — **1 122 320 384 bytes** (около 1,045 GiB);
- **0** затронутых range-разбиений.

template0 не измерена: подключение фактически отклонено. Ноль относится только к 30 измеренным базам. `template0` не квалифицирована как источник клонирования; без отдельной будущей qualification её нельзя использовать как проверенный cloning source или включать в заявление о полной готовности recovery. Открывать её или менять `dataallowconn` ради настоящего документа нельзя.

Время REINDEX и численный RTO не измерены. Только сценарная чувствительность компонента по объёму: V / v_{eff} , где $V = 1\,070,328125$ MiB. При условных 10 / 50 / 100 MiB/s это примерно 107,03 / 21,41 / 10,70 секунды. Эти скорости — предположения, не замер данного host и не обещанный диапазон производительности. Полная добавка к RTO включает чтение hear, сортировку, WAL, конкуренцию за ресурсы и накладные расходы; они не измерены.

Условная процедура recovery после promotion

Приложения не подключать на запись до завершения соответствующей ветки и проверки корректности. Процедура ниже — документ будущей разрешённой операции, не команда исполнять её сейчас.

0. Promotion и актуальная проверка range partitions

После разрешённого `pg_promote()` / `pg_ctl promote` в каждой базе повторить запрос ниже. Использовать актуальное состояние на момент аварии, а не прошлый нулевой снимок NGw4. `template1` входит в проверку; предел `template0` сохраняется отдельно.

```
SELECT n.nspname, c.relname, pg_get_partkeydef(c.oid) AS partkey
FROM pg_partitioned_table p
JOIN pg_class c ON c.oid = p.partrelid
JOIN pg_namespace n ON n.oid = c.relnamespace
WHERE p.partstrat = 'r'
AND EXISTS (
  SELECT 1
  FROM unnest(p.partcollation::oid[]) AS k(coll)
  JOIN pg_collation co ON co.oid = k.coll
  JOIN pg_database d ON d.datname = current_database()
  WHERE (co.collprovider = 'c' AND co.collcollate NOT IN ('C', 'POSIX'))
        OR co.collprovider = 'i'
        OR (co.collprovider = 'd' AND (d.datlocprovider = 'i'
        OR (d.datlocprovider = 'c' AND d.datcollate NOT IN ('C',
'POSIX'))));
```

Не-collatable ключи с `oid 0` отсекаются `join`; builtin и C/POSIX не выбираются. List/hash-разбиение не входит в принятую проверку порядка при детерминированной collation.

1. Непустая выборка — остановка перед записью

Если найден хотя бы один range-ключ с затронутой collation, этот recovery-рецепт недостаточен. Приложения на запись не подключать.

REINDEX не перемещает строки между разделами. До включения записи требуется проверить/исправить размещение строк найденных таблиц либо получить решение о совместимом target. Список таблиц немедленно передать OLVR/Борису. Проектирование repartitioning этим runbook не выполняется; перенос на Linux не разрешён решением владельца.

2. Пустая выборка — перестроение индексов

Если актуальная выборка пуста во всех проверенных базах, в каждой базе, включая `template1`, выполнить REINDEX индексов с затронутой collation. Принятый контракт допускает REINDEX DATABASE для соответствующей базы.

Не выдавать это условие за выполненное для `template0`: её техническая граница требует

отдельного disposition и будущей qualification. Не объявлять полную готовность всех баз при сохраняющемся пробеле.

3. Обновление metadata collation

После перестроения выполнить для каждой соответствующей базы:

```
ALTER DATABASE <db> REFRESH COLLATION VERSION;
```

<db> — placeholder фактического, корректно экранированного имени базы, а не буквальная команда. REFRESH сам по себе не исправляет индексы или range-размещение.

4. Проверка корректности

В рамках той же разрешённой recovery-операции: CREATE EXTENSION IF NOT EXISTS amcheck, затем bt_index_check по индексам с затронутой collation — **без ошибок**. Это будущие изменяющие/проверочные действия; сейчас они не выполнялись.

Проверка отсутствия WARNING при новом подключении касается только metadata. Она не заменяет amcheck и не снимает ветку остановки при range partitions.

5. Подключение приложений

Только после необходимого REINDEX, REFRESH и успешной проверки корректности подключать приложения на запись. При непустой range-выборке сначала обязательно завершить ветку проверки/исправления размещения либо получить решение о совместимом target. Непроверенную template0 не включать в обещание полной готовности или безопасного клонирования.

Решение владельца и дальнейшая ответственность

Борис через OLVR, Yqlv/IFILCU: «**4416 — никаких линуксов не будет**». Перенос standby на Linux исключён. **Переход на ICU не разрешён.**

Отказ от Linux не означает устранения mismatch, приемлемого RTO или готовности горячей замены. Долгосрочное решение и технический disposition template0 остаются за OLVR/Борисом в OLV-4416. Jira пока не закрывать как устранённую.

Риск той же природы возникнет на primary при будущей смене ОС/glibc srv-prod-postgre.olvr. До такого изменения учитывать зависимость индексов и range-размещения от правил collation.

Основания

- Принятый DESIGN: YqIv/gvQSwN; независимая приёмка YqIv/NGmhXR.
- Принятые измерения: NGw4/9jyCrC; полный измерительный handoff NGw4/AjuxDx.
- Решение владельца: YqIv/IFILCU; документационное поручение k4Bf/HSOJfN.
- PostgreSQL 17: [ALTER COLLATION](#); [Locale data changes](#).

Публикация этого runbook — документационный результат, не выполнение recovery и не устранение collation mismatch.

From:

<https://wiki.olvery.ru/> - **Olvery**

Permanent link:

<https://wiki.olvery.ru/doku.php?id=runbooks:srv-postgre-olvr>

Last update: **2026/10/02 11:54**

