

ozon{tech

ozon{ech

Опыт эксплуатации pg_probascur на 1.5k ИНСТАНСОВ

Смолкин Григорий
инженер

Система резервного копирования

Формируется:

- Требованиями бизнеса
- Нюансами конкретной экосистемы



Ретроспектива

Требования бизнеса год назад

- RPO: 1d
- RTO: 1-2h
- Retention: 7d

Экосистема год назад

- Ansible
- Patroni
- ~500 инстансов PostgreSQL
- Общий объем* PostgreSQL: 28TB

Наше решение год назад

- pg_probackup via ssh + **Ansible**
- автономные бэкапы (--stream)
- x2 железный сервер под хранилище (73TB)
- incremental ratio: ~0.12*
- avg compression ratio: ~3.13
- retention: 2w



Новое требование бизнеса

Production to Staging

Production to Staging

- внедряем **incremental restore**
- avg incremental ratio: **~22%**

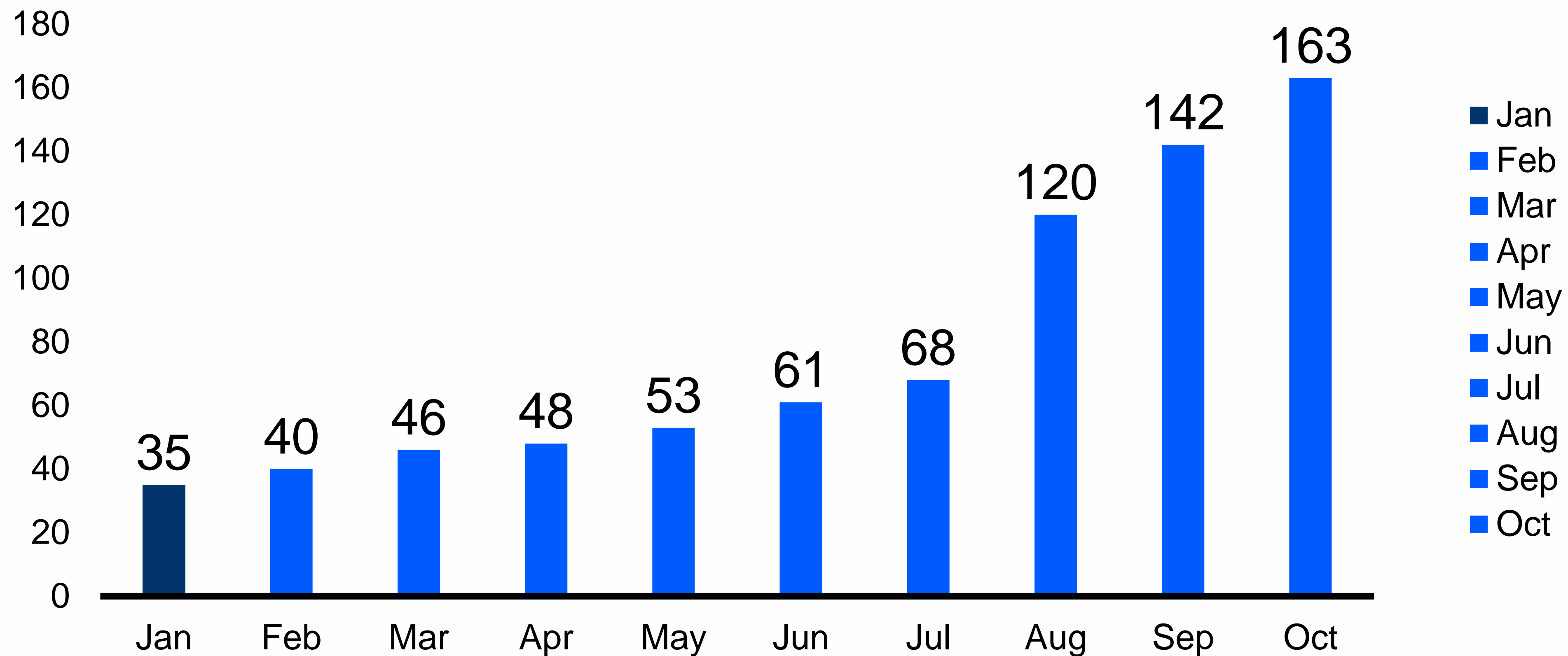


Production растёт

Перестаем влезать в 73ТВ

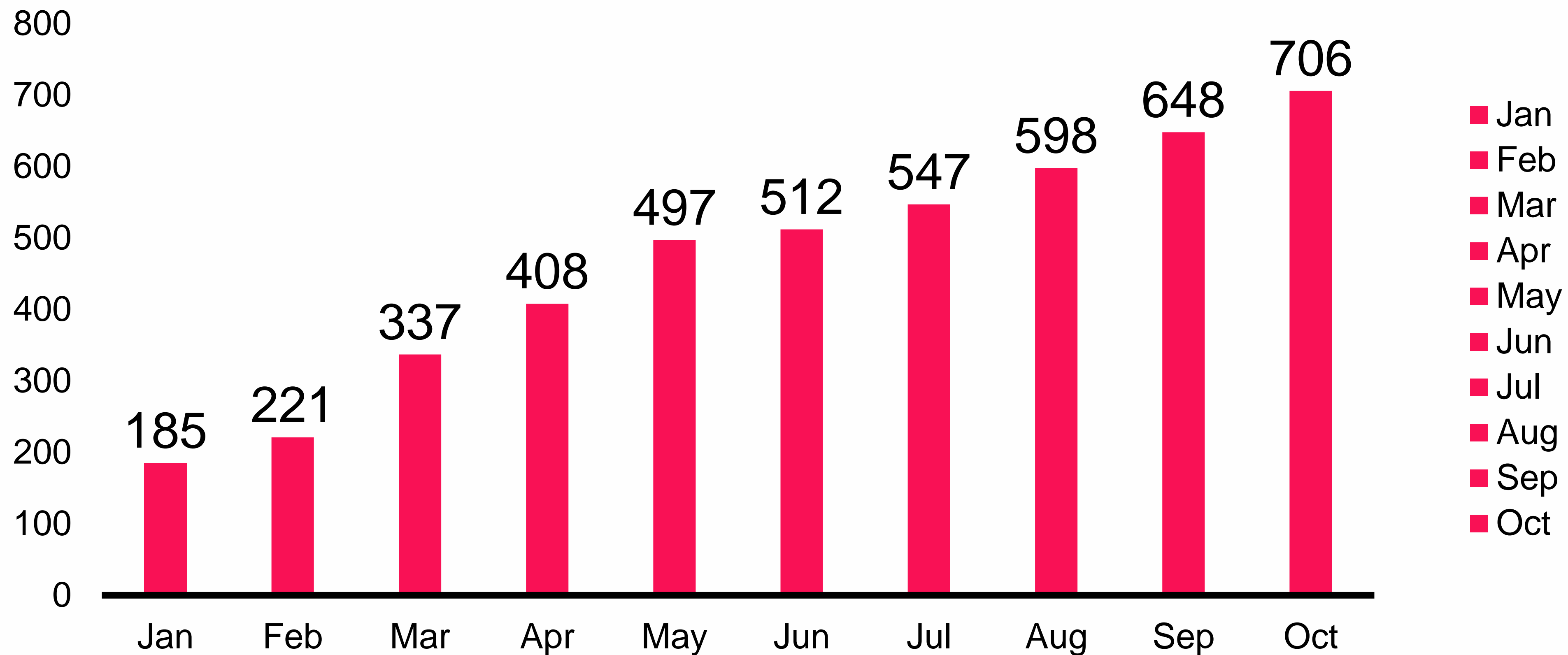
Production растёт

Суммарный размер всех БД в ТВ (masters only)



Production растёт

Суммарная скорость генерации WAL в MB



Затягиваем пояса

- включаем `--no-validate`
- включаем `--no-sync`
- внедряем `merge retention`
- переходим на DELTA-only стратегию

FULL -> Delta-> Delta->Delta->Delta-> Delta-> **FULL**->Delta->Delta->Delta->Delta->Delta->Delta->Delta->**FULL**

FULL -> Delta-> Delta->Delta->Delta->Delta->Delta->Delta->Delta-> Delta-> Delta->Delta

Затягиваем пояса

- включаем `--no-validate`
- включаем `--no-sync`
- внедряем `merge retention`
- переходим на DELTA-only стратегию

FULL -> Delta-> Delta->Delta-> Delta-> Delta->Delta -> **FULL** -> Delta-> Delta->Delta-> Delta-> Delta->Delta

FULL -> Delta-> Delta->Delta-> Delta-> Delta->Delta -> Delta-> Delta->Delta-> Delta-> Delta->Delta

Улучшаем pg_probackup

1. merge `--no-validate`
 2. merge `--no-sync`
 3. restore `--force`
 4. перенос fsync WAL в конец процесса бэкапа
 5. учимся детектить timeline switch
 6. стабильное API ssh агента
 7. улучшения блокировок резервных копий
 8. множество других мелких улучшений и фиксов
- TODO: куча идей (direct I/O, WAL stream compression, merge retention-bytes)



Новое требование бизнеса

Backup to Sandbox



Новое требование бизнеса

Еженедельное восстановление резервной копии для выбранного числа баз

Последний туз в рукаве

- На бэкапной машине становится жарковато
- Инкрементальное восстановление спасает
- Обрезаем retention до 7d

CALM DOWN BRO





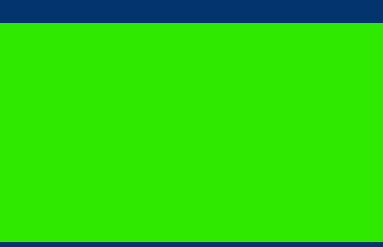
Новое требование бизнеса

MULTI-DC



Новое требование бизнеса

Point-In-Time-Recovery



Новое требование бизнеса

Retention: user-defined

Запас фокусов исчерпан

Надо менять архитектуру системы
резервного копирования



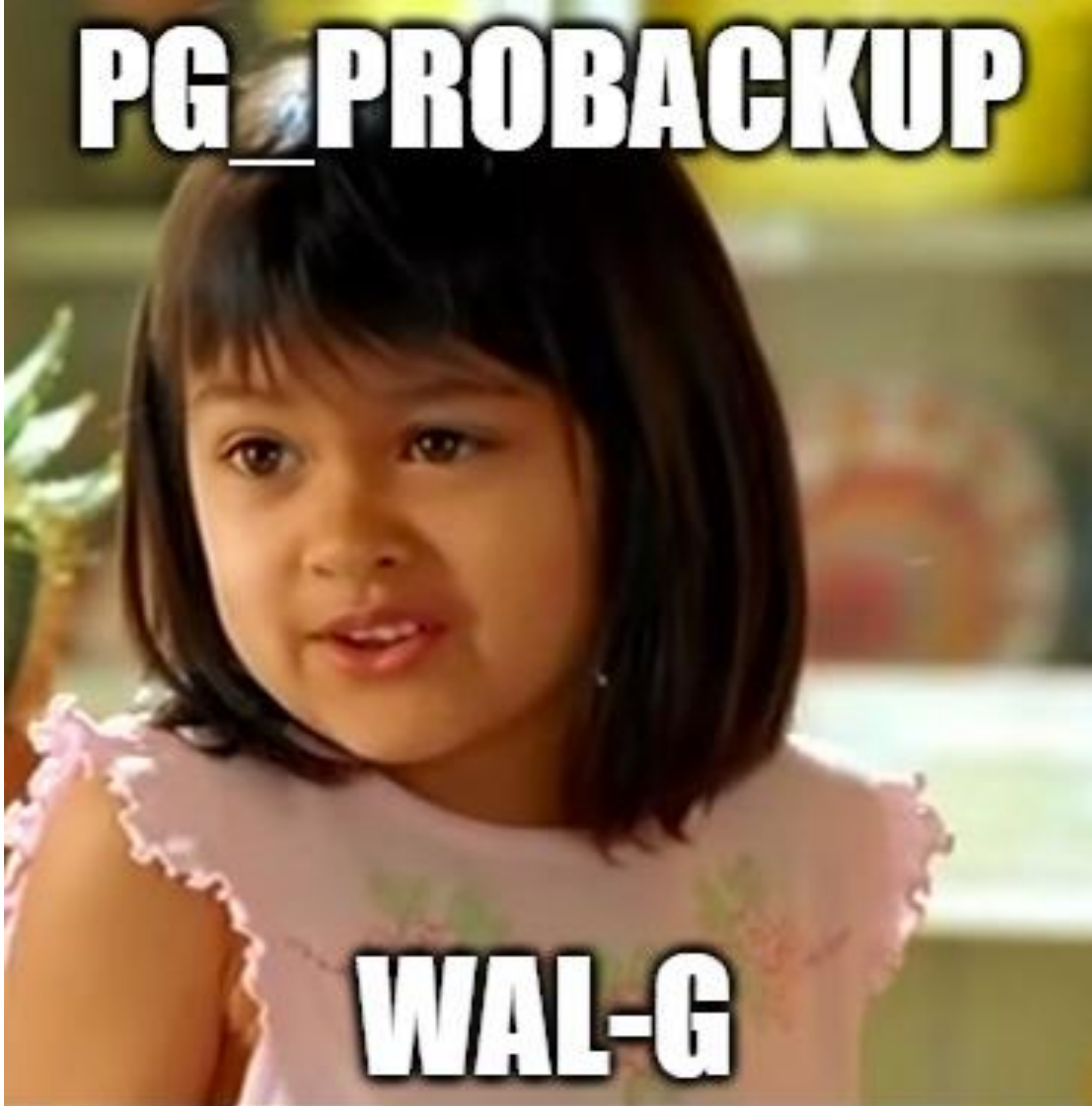
Требования к стораджу

- избыточность
- масштабируемость
- изоляция
- склонность к самолечению
- быть головной болью каких-то других людей

Муки выбора

1. pg_probackup + пачка железных машин
2. wal-g + ceph

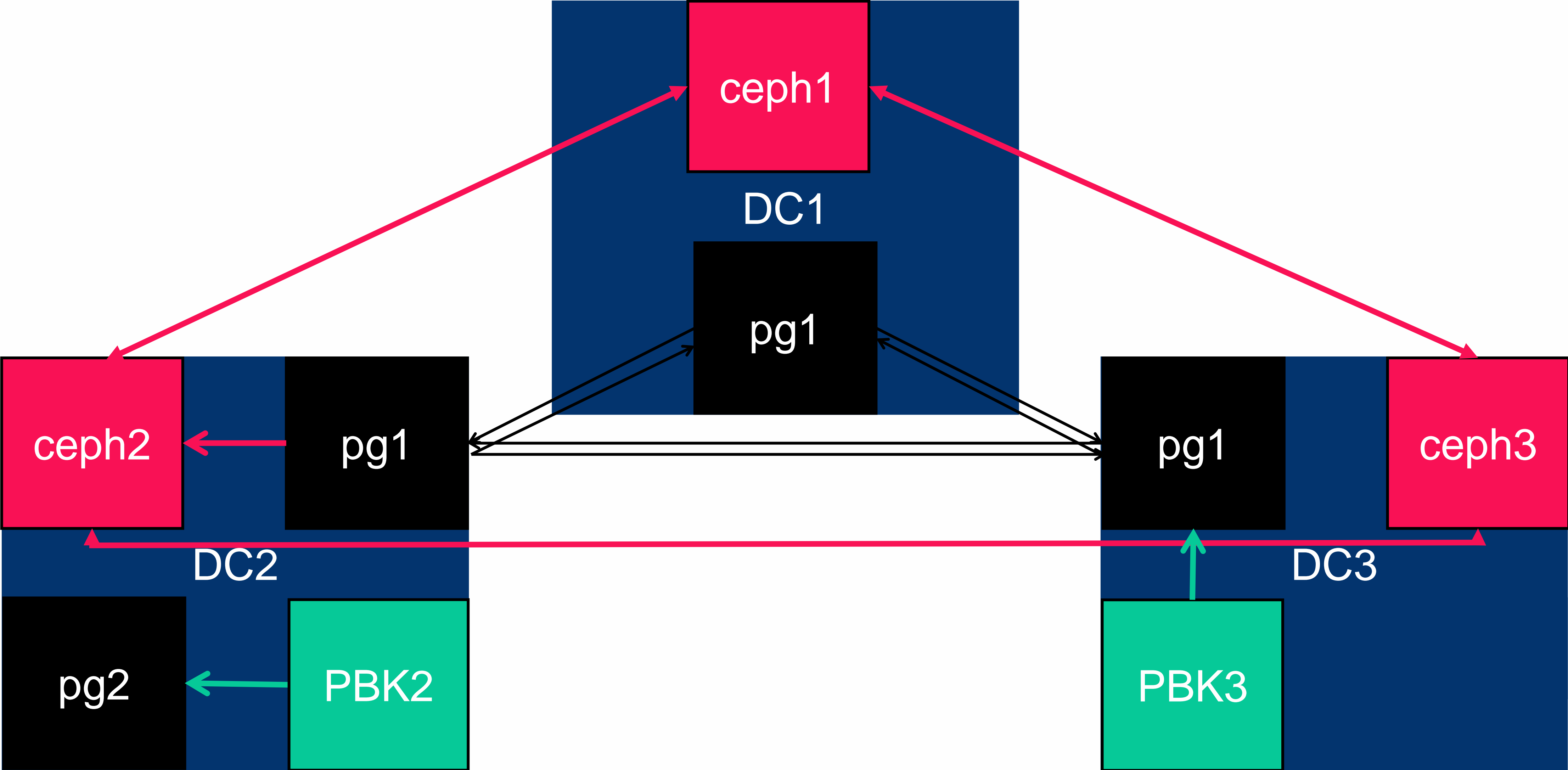
PG_PROBACKUP



WAL-G

Why don't we have both?

Текущая схема бэкапирования



В планах

1. patroni + pg_probackup **catchup**
2. pg_probackup + wal-g = love
 - файлы данных приносит pg_probackup
 - WAL приносит wal-g
3. S3 for pg_probackup ?

ozon{ech

Спасибо за внимание!

Смолкин Григорий



gsmolkin@ozon.ru, g.smolkin@postgrespro.ru