

# Arenadata DB

Сегодня и завтра



Head of Software Architecture  
Дмитрий Воронков



## План

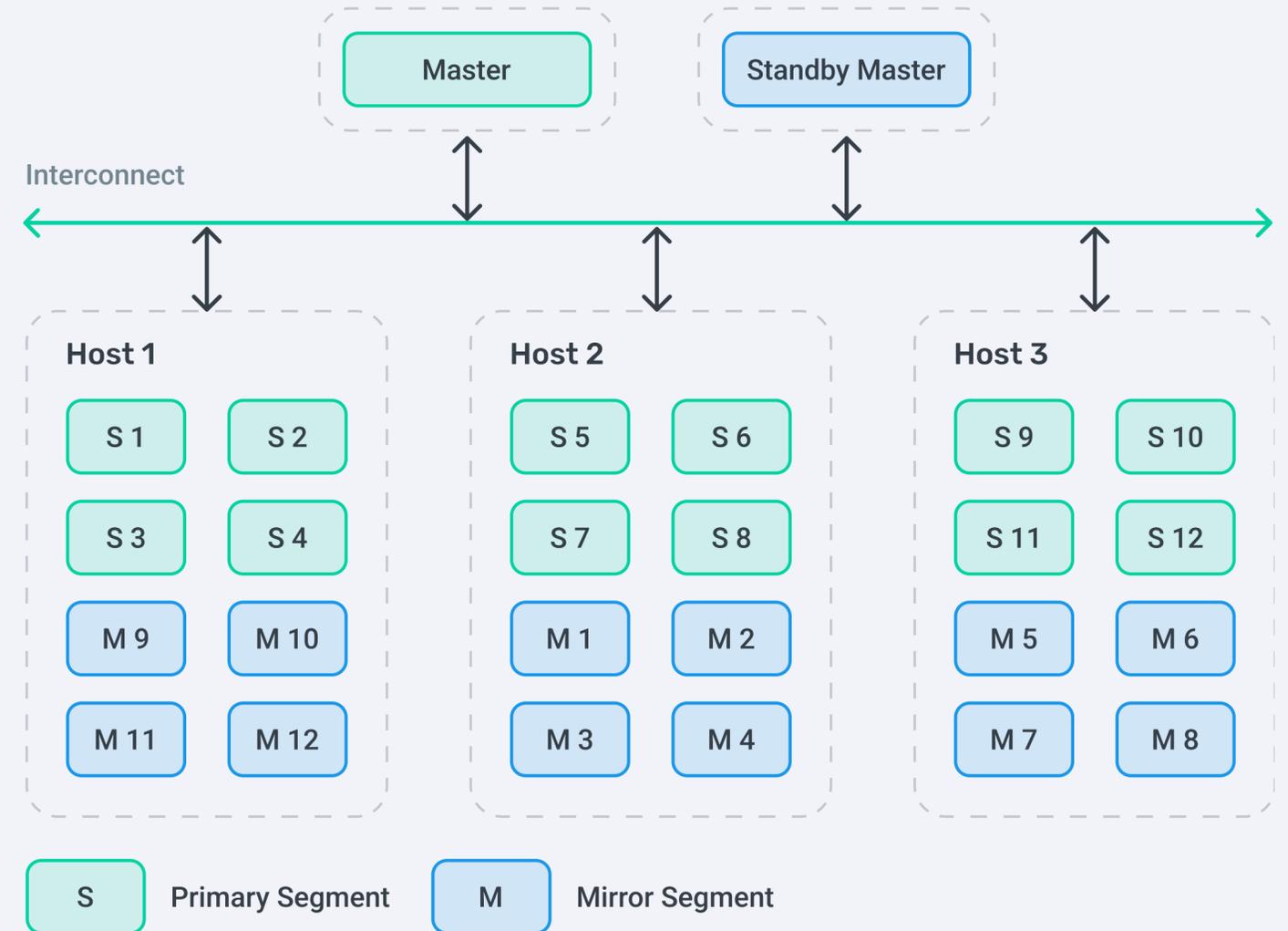
- 1 Продукт Arenadata DB
- 2 Наш вклад в Open Source
- 3 Возможности Enterprise Edition
- 4 Мониторинг запросов
- 5 Управление резервными копиями
- 6 Новости по ближайшим релизам
- 7 Планы по Greenplum 7
- 8 Планы по развитию своего форка
- 9 Вопросы и ответы



Arenadata DB

## Arenadata DB

Массивно-параллельная реляционная СУБД с открытым исходным кодом для хранилищ данных с гибкой горизонтальной масштабируемостью и колоночным хранением на основе PostgreSQL.



## Arenadata DB: ключевые особенности



Производительность



Безопасность



Надежность



Удобство развертывания  
и конфигурирования



Контрибуция в Greenplum



Arenadata DB

Вклад в Open Source

## Arenadata DB в цифрах

**10 релизов**

Выпущено за последний год



**7 из них**

Выпущены с обновлением  
версии Greenplum

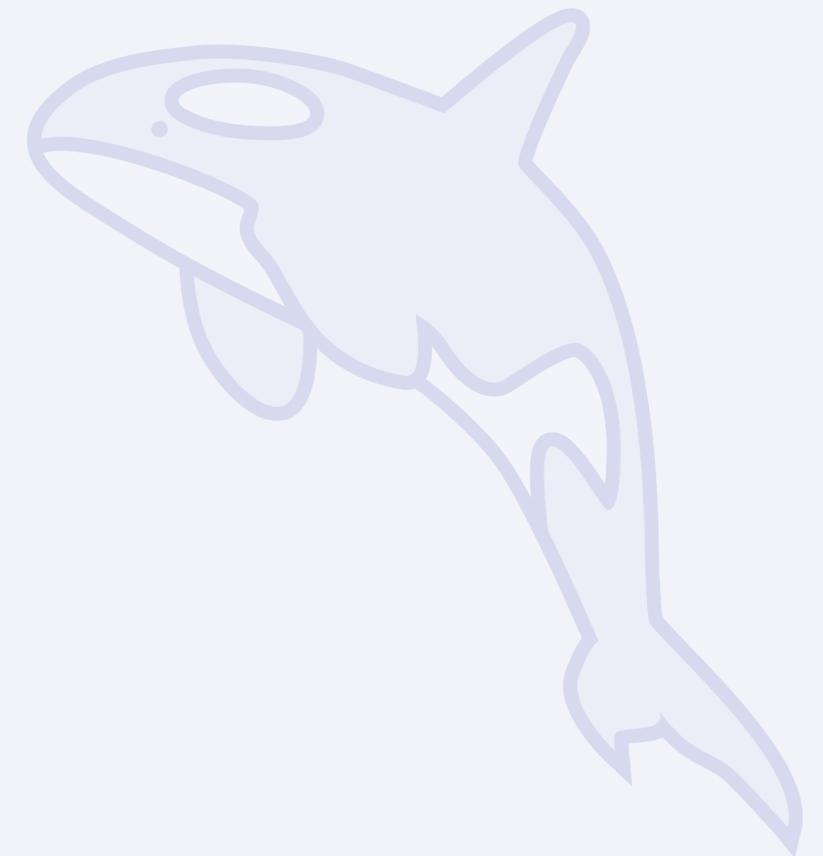


## Arenadata DB

## Вклад в Open Source

## Arenadata DB за последний год. СУБД Contribution

- ✓ Доработки планировщика GPORCA
- ✓ Рефакторинг исходников Greenplum 7
- ✓ Переработка функции смены ресурсной группы для исполняемого запроса;
- ✓ Доработка расширений:
  - diskquota;
  - gp\_toolkit.
- ✓ Улучшение утилит по работе с логическими резервными копиями:
  - gpbackup;
  - gprestore.



Arenadata DB

Вклад в Open Source

Enterprise Edition

## Arenadata DB Enterprise Edition

Помимо СУБД, Arenadata DB Enterprise Edition поставляется с дополнительными компонентами, предоставляющими решение для различного рода задач.



Arenadata DB

Вклад в Open Source

Enterprise Edition

## Arenadata DB за последний год

- ✓ Расширение функционала arenadata\_toolkit
- ✓ Добавлена поддержка AstraLinux 1.7 SE "Orel"
- ✓ Исправления в источниках на основе статического анализатора
- ✓ Обновление Greenplum до версии 6.26
- ✓ Добавлена поддержка plpython3u;

- ✓ diskquota
- ✓ Parameter archiving
- ✓ gpbackup/gprestore:
  - доведение до рабочего состояния опции восстановления на другое число сегментов;
  - оптимизация времени восстановления при использовании плагина adb\_ddp\_plugin.

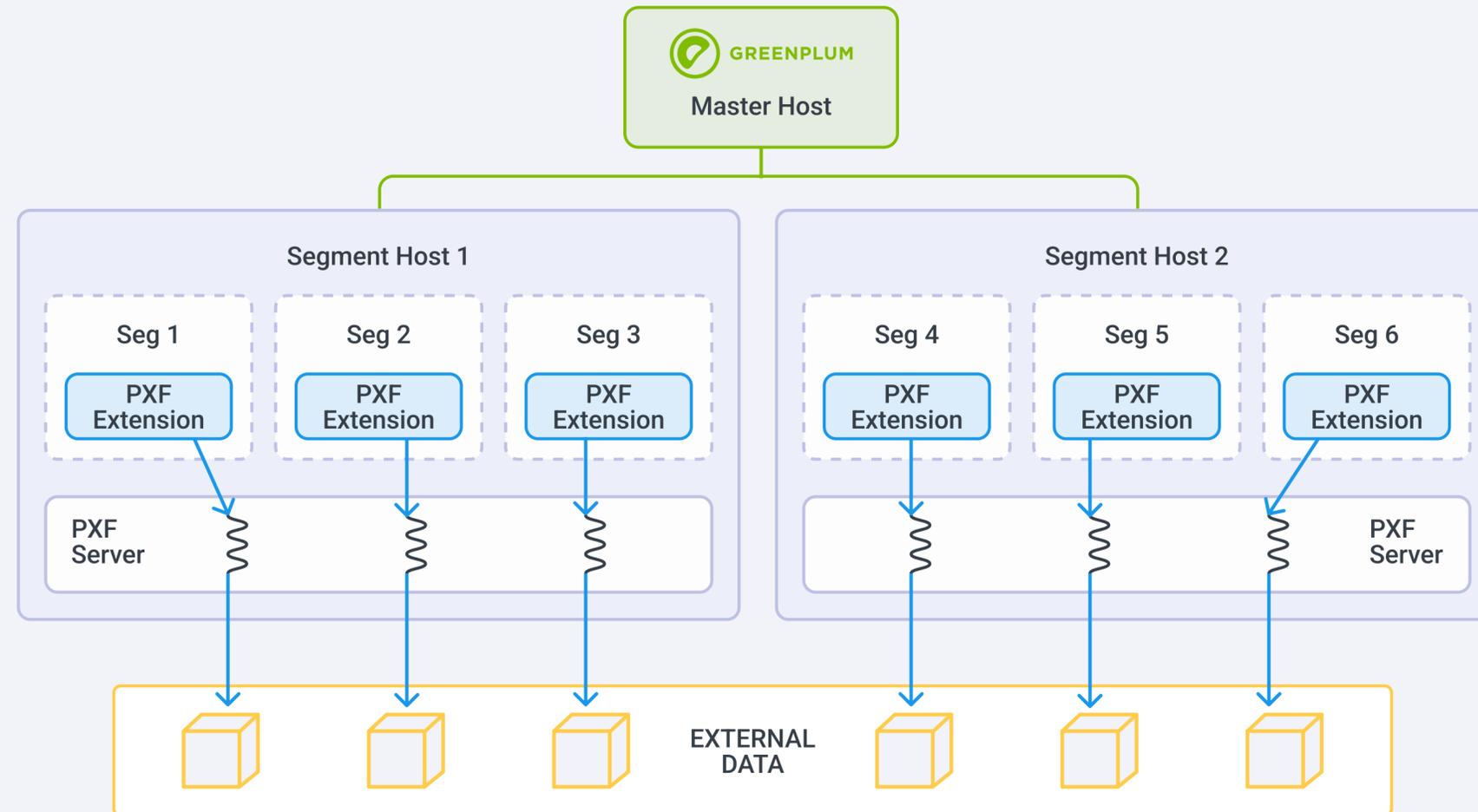
Arenadata DB

Вклад в Open Source

Enterprise Edition

## PXF Connector

Фреймворк для параллельного и высокопроизводительного доступа к гетерогенным источникам данных из **Arenadata DB** при помощи встроенных коннекторов.



Arenadata DB

Вклад в Open Source

Enterprise Edition

## PXF Connector

- Поддержка **push-down** для предиката **IN** в **PXF JDBC**;
- реализовано корректное прерывание читающих запросов в **PXF JDBC**;
- добавлена команда перезагрузки профилей;
- расширен список метрик сервиса **PXF**;
- добавлены дашборды для мониторинга сервиса **PXF** в **Grafana**.



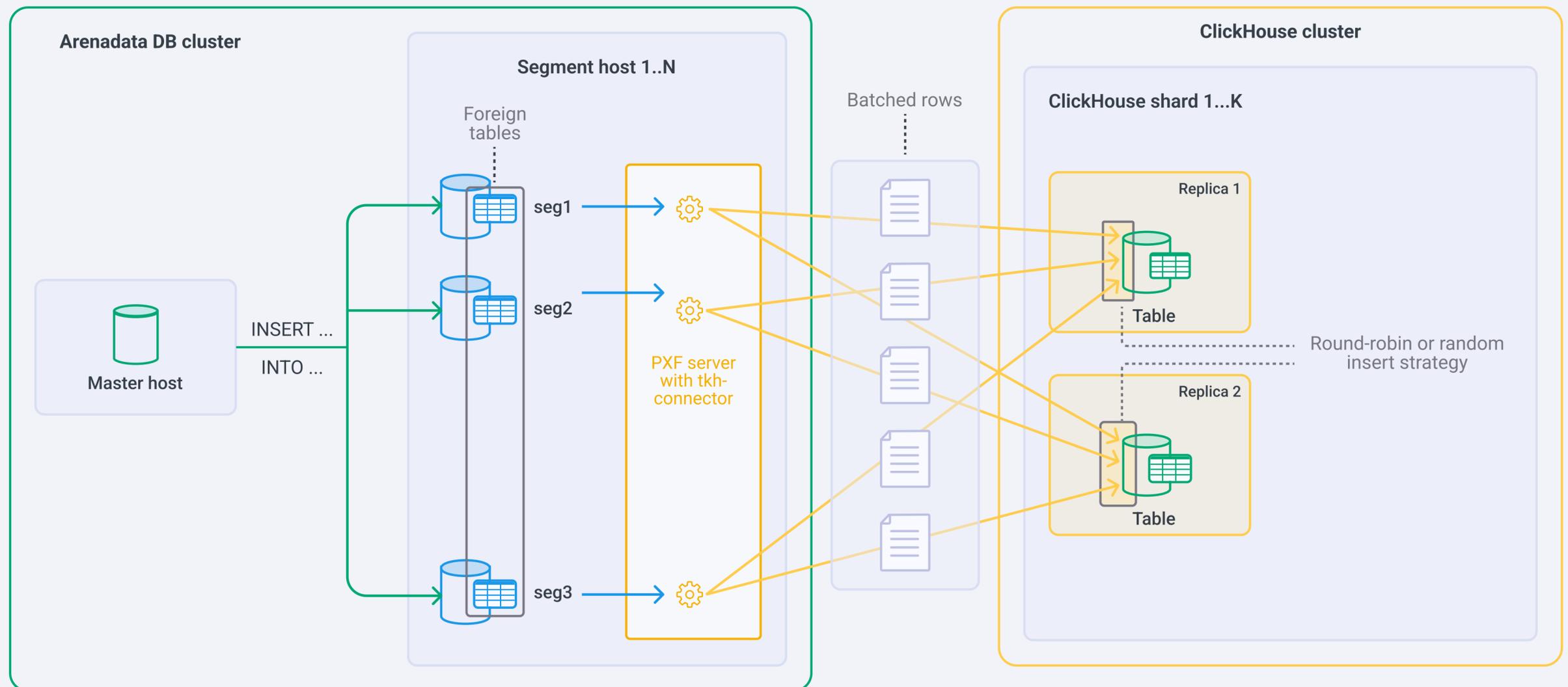
Arenadata DB

Вклад в Open Source

Enterprise Edition

## ADB Clickhouse Connector

FDW-коннектор (foreign data wrapper) для передачи данных в Arenadata QuickMarts или ClickHouse.



Arenadata DB

Вклад в Open Source

Enterprise Edition

## ADB Clickhouse Connector

### ✓ Релиз Tkhemali 2.0

- Транзакционно загружает данные за счет автоматического создания staging-таблиц;
- использует несколько семейств engine таблиц в Clickhouse;
- гибко распределяет и распараллеливает нагрузку записи.



[Arenadata DB](#)

[Вклад в Open Source](#)

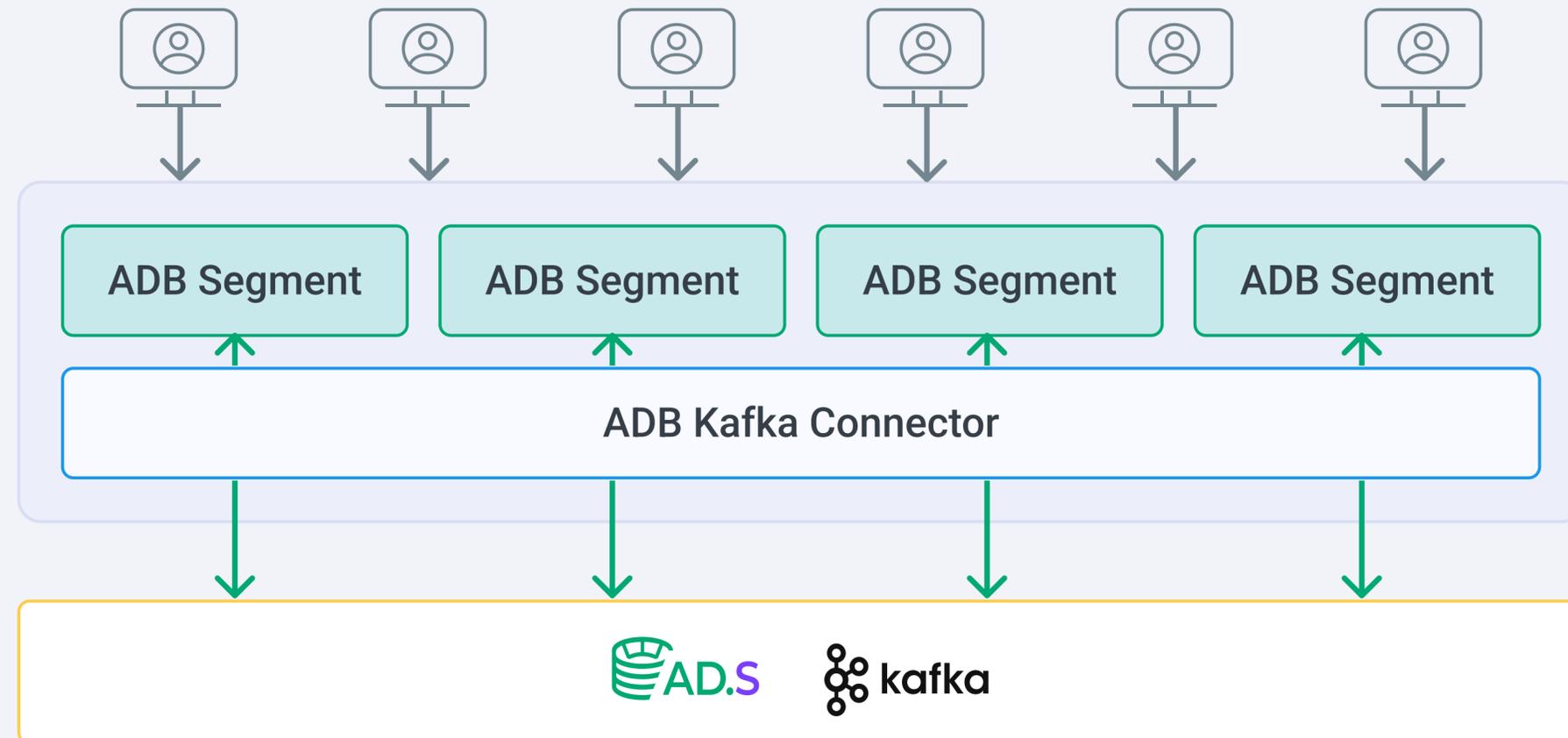
[Enterprise Edition](#)

Новости

## ADB Kafka Connector

Специализированный коннектор для интеграции **Apache Kafka** с **Arenadata DB**.

- Добавлена поддержка **SASL\_PLAINTEXT** и **SASL\_SSL** для настройки аутентификации.



Arenadata DB

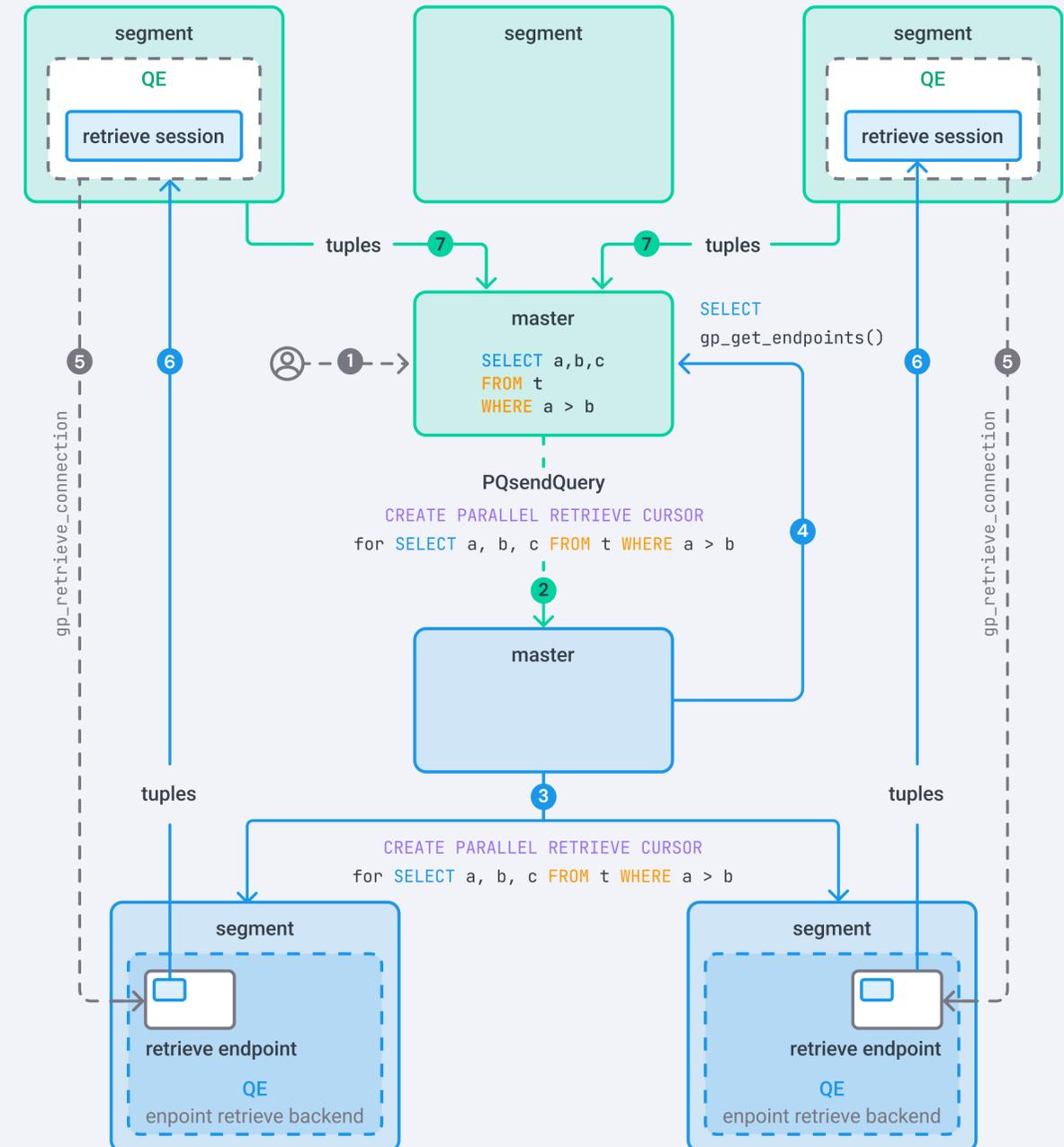
Вклад в Open Source

Enterprise Edition

## ADB to ADB

В релиз вошли следующие возможности:

- параллельное получение данных сегментами локального кластера напрямую с сегментов удаленного кластера;
- транзакционная вставка данных с локального на удалённый кластер в режиме «master-master»;
- возможность автоматического определения количества обработчиков.



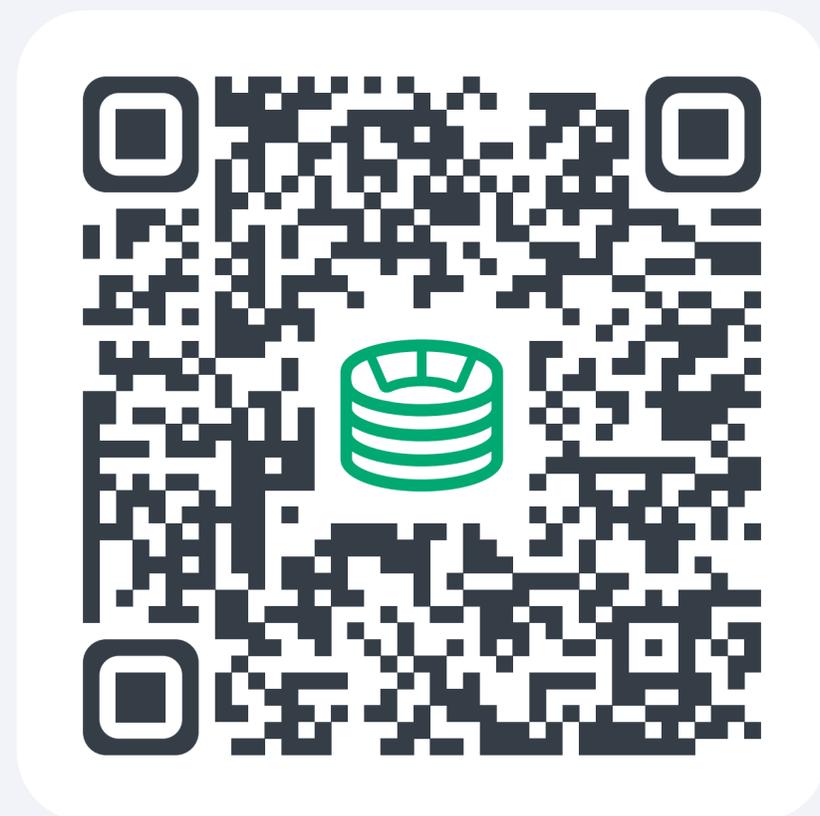
Arenadata DB

Вклад в Open Source

Enterprise Edition

# Коннектор ADB-TO-ADB

[habr.com/ru/companies/arenadata/](https://habr.com/ru/companies/arenadata/)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Система мониторинга запросов Arenadata DB в режиме реального времени.

Предназначена для глубокого исследования процессов выполнения команд или утилит, работающих на кластерах ADB.

[Обзор Arenadata DB Control](#)

The screenshot displays the AD.B CONTROL web interface. At the top, it shows the date and time (04/02/2022 14:49:36) and the language (EN). The main navigation bar includes 'Monitoring', 'Commands', 'Transactions', 'Sessions', and 'Resource groups', with 'Commands' currently selected. Below this, there are tabs for 'Online' and 'History', with 'Online' active. A status indicator shows 'running 1'. The 'Cluster' is set to 'Production, UAT'. A table lists active commands with columns for Command ID, Command text, SQL ID, and Username. The table contains one entry with a green play button icon.

Command ID	Command text	SQL ID	Username
1709568945-571373-14-1753018553	insert into public.t_big_table(x, txt1, txt2, txt3) select i, md5(i:text...	-908512937	gpadmin

[Arenadata DB](#)[Вклад в Open Source](#)[Enterprise Edition](#)[Мониторинг запросов](#)

Основные функции, реализованные в прошедших релизах:

- ✓ Выгрузка метрик во внешнее хранилище для долгосрочного хранения и построения аналитики
- ✓ Задачи ИБ:
  - CSRF, XSS protection;
  - Управление паролями;
  - Аудит действий;
  - Поддержка SSL/TLS между сервисами;
- ✓ Улучшена интеграция через LDAP
- ✓ Управление GUC из UI, конфигурирующих мониторинг
- ✓ Расширение функционала мониторинга запросов:
  - мониторинг и прерывание сессий;
  - добавлен `sql_id` для запросов;
  - маркировка запросов при помощи `query_band`;
  - расчет **total метрик** на лету;
  - мониторинг запросов на основе продолжительности выполнения;
- ✓ Поддержка high availability для внешнего кластера Clickhouse;



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

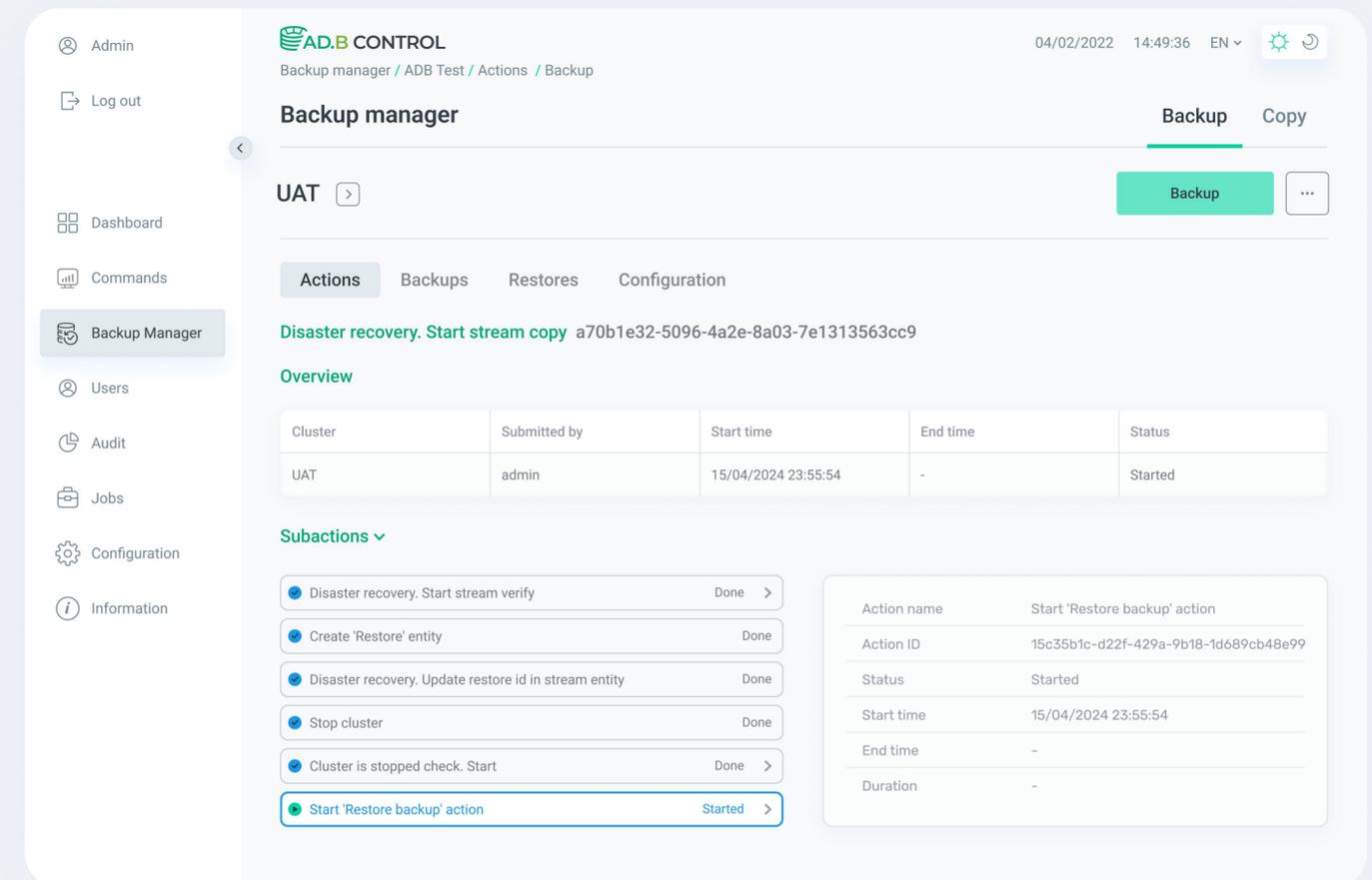


Сервис по управлению бинарными бэкапами для кластеров ADB.

Основные улучшения:

- ✓ Восстановление на кластере с другой конфигурацией топологии;
- ✓ Расширение опций восстановления:
  - delta restore;
  - force restore;
  - параллельное и последовательное восстановление зеркал;
- ✓ Поддержка CIFS репозитория;
- ✓ Переработаны механизмы работы с репозиториями и очистки данных.

[Обзор Arenadata DB Backup Manager](#)



Arenadata DB

Вклад в Open Source

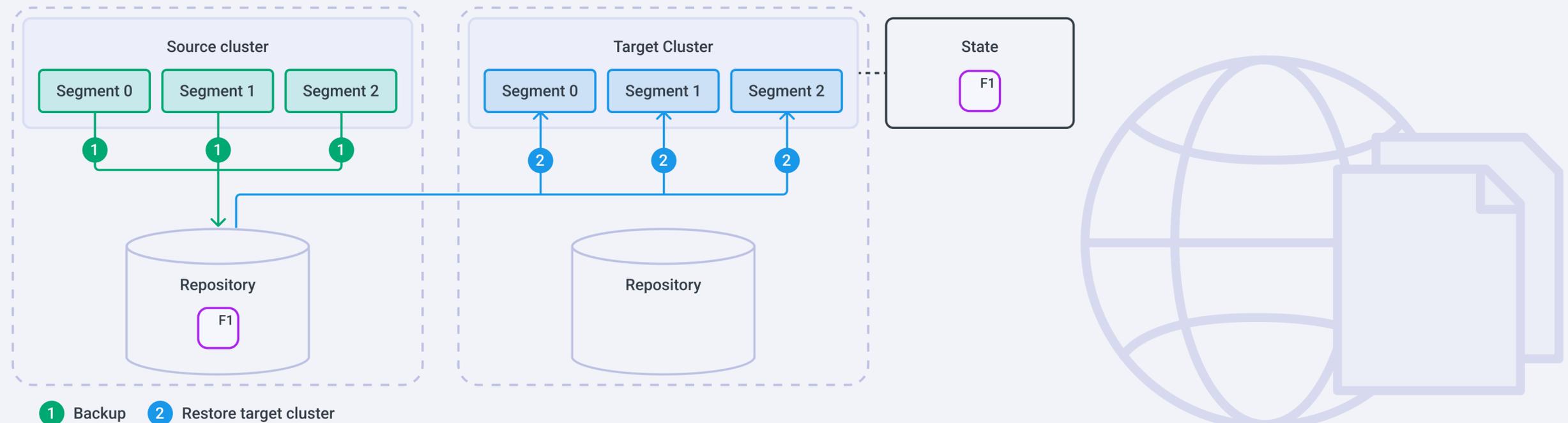
Enterprise Edition

Мониторинг запросов

Управление бэкапами

Реализовано решение для сценария Disaster Recovery в нескольких вариантах:

- ✓ Cold Standby через восстановление из бэкапа другого кластера
- ✓ Автоматическое применение бэкапа на удаленном кластере
- ✓ Warm Standby на базе стриминга изменений на удаленный кластер между точками восстановления
- ✓ Проверка доступности репозитория
- ✓ Сохранение топологии целевого кластера после восстановления



Arenadata DB

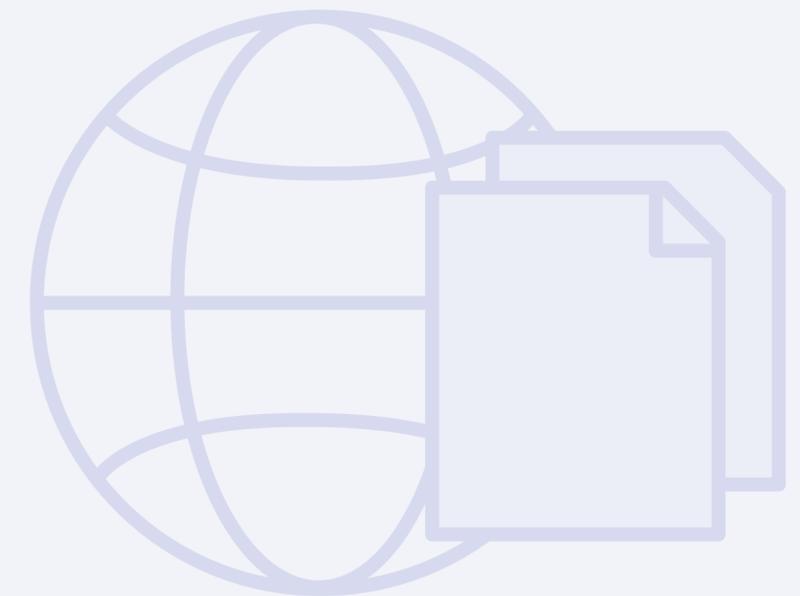
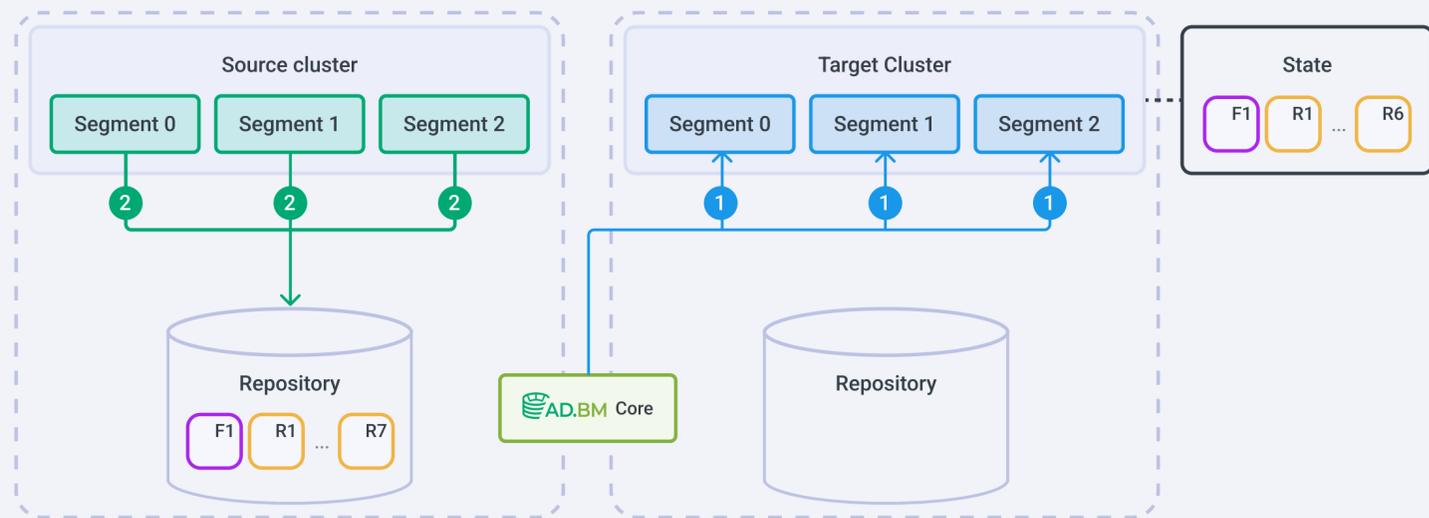
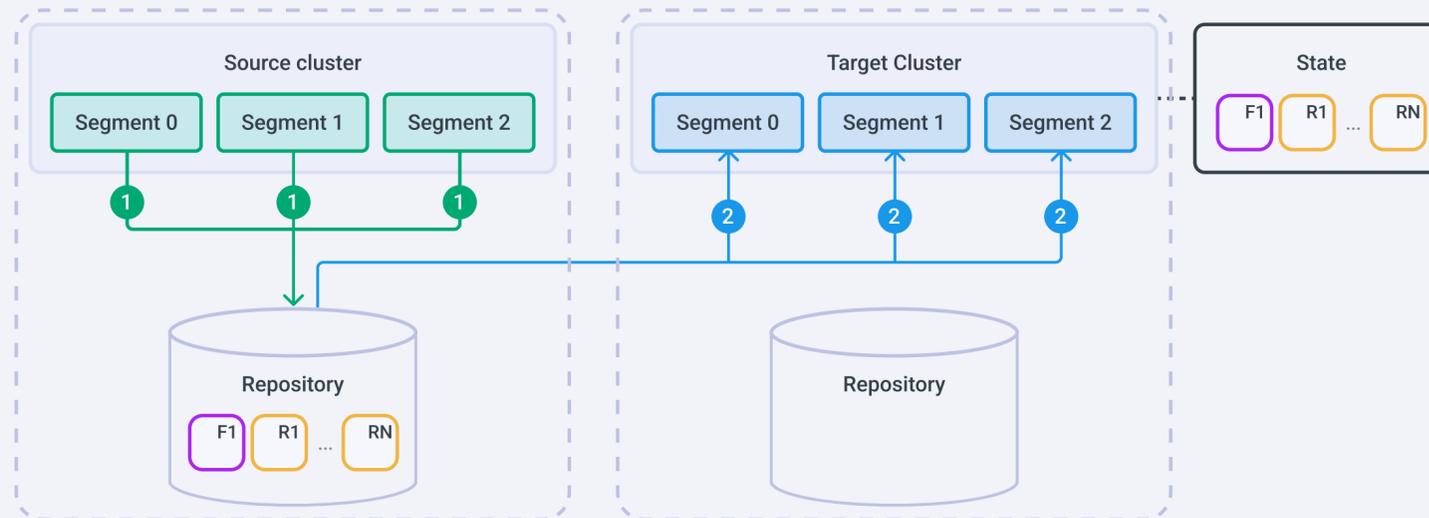
Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

## AD.BM Disaster Recovery



[Arenadata DB](#)[Вклад в Open Source](#)[Enterprise Edition](#)[Мониторинг запросов](#)[Управление бэкапами](#)[Новости](#)

## Новости по ближайшим релизам

### ✓ Arenadata DB

- Поддержка **Ubuntu 22.04**;
- Трансформации по умолчанию для **gpfdist**;
- Безопасные настройки доступа в **pg\_hba.conf** и включение **ident server** в поставку;
- Валидация конфигурации **pg\_hba** до применения на кластере;
- Разделение хранения данных временных таблиц и вспомогательных файлов;
- Выпуск обновленной CE версии на базе **Greenplum 7.1**.

### ✓ ADB Control

- Интеграция с хранилищем секретов;
- Поддержка **LDAPs**;
- Задачи **ИБ**.

### ✓ Backup Manager

- **CLI** утилита по управлению логическими резервными копиями;
- Поддержка работы с кластером после **expand**.

### ✓ PXF

- Добавление метрик мониторинга пула потоков;
- Расширение списка поддерживаемых типов данных для **push-down** фильтрации.

[Arenadata DB](#)[Вклад в Open Source](#)[Enterprise Edition](#)[Мониторинг запросов](#)[Управление бэкапами](#)[Новости](#)[Планы по Greenplum 7](#)

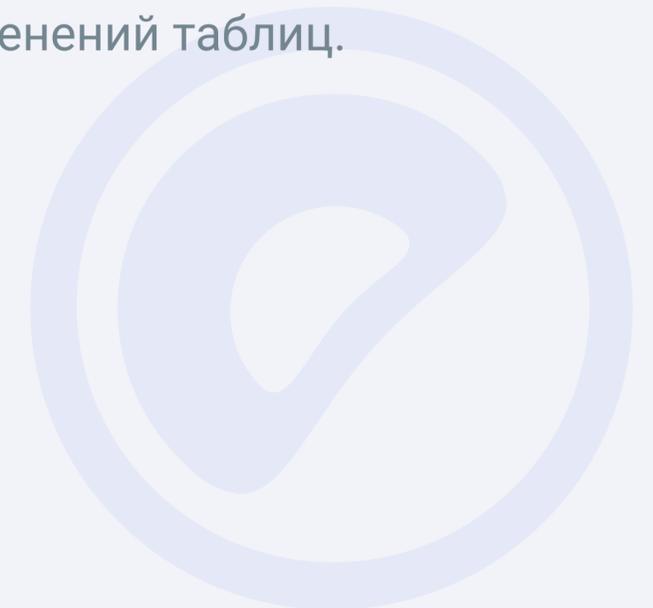
## Планы по Greenplum 7

### ✓ ADB Community Edition

- Реализация работы с партицированными таблицами в расширениях;
- Переработка структуры бандла;
- Работа по поддержке **cgroup2**;
- Изменение механизма запуска **gpexpand** из **ADCM**;
- Стабилизация версии.

### ✓ ADB Enterprise Edition

- Доработка расширений;
- Развитие коннектора **ADB to ADB** и исследование по расширению функционала (**push-down** агрегатов и соединений);
- Портирование коннектора **ADB Clickhouse**;
- Сделать **fdw-kafka** коннектор двунаправленным;
- Отслеживание изменений таблиц.



[Arenadata DB](#)

[Вклад в Open Source](#)

[Enterprise Edition](#)

[Мониторинг запросов](#)

[Управление бэкапами](#)

[Новости](#)

**Планы по Greenplum 7**

## Планы по Greenplum 7

### ✓ ADB Control

- Поставка в виде отдельного бандла;
- Мониторинг ожиданий;
- Сервис исторических данных;
- Проверки на основе правил и реакции.

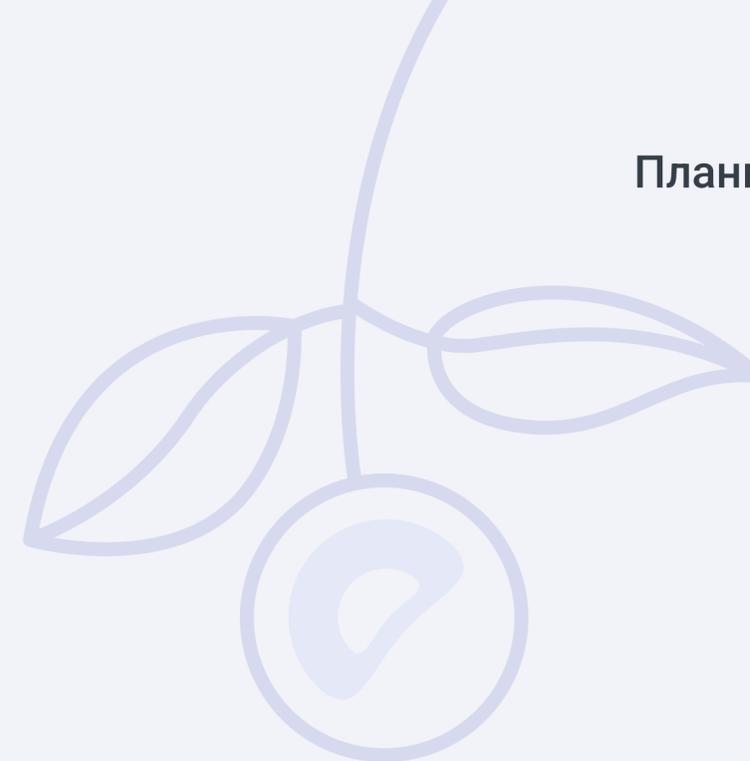
### ✓ Backup Manager

- Управление логическими резервными копиями и их конфигурацией из **UI**;
- Запись в несколько репозиториев;
- Восстановление отдельных таблиц из бинарной резервной копии.



## Что нового в Greenplum 7 (часть 1)

[habr.com/ru/companies/arenadata/](https://habr.com/ru/companies/arenadata/)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

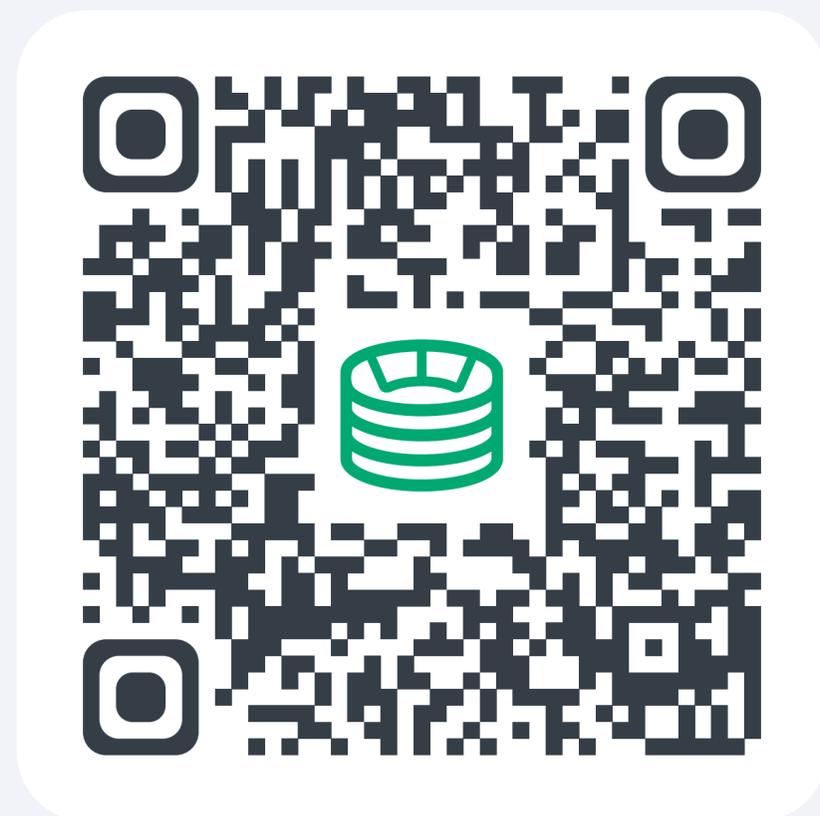
Управление бэкапами

Новости

Планы по Greenplum 7

## Что нового в Greenplum 7 (часть 2)

[habr.com/ru/companies/arenadata/](https://habr.com/ru/companies/arenadata/)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

Новости

Планы по Greenplum 7

## Далее

- ✓ Что нового в Greenplum 7 (часть 3)
- ✓ Отладка JIT в Greenplum/Postgres
- ✓ Как Postgres инициализирует базу данных
- ✓ Free Space Map в Postgres

[habr.com/ru/companies/arenadata/](https://habr.com/ru/companies/arenadata/)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

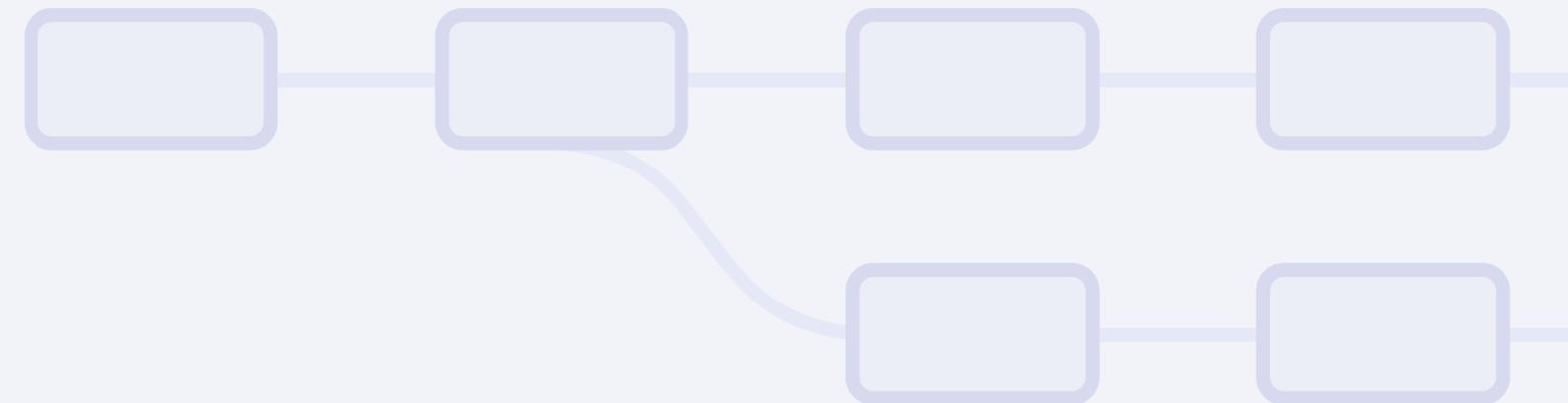
Новости

Планы по Greenplum 7

Планы по развитию  
своего форка

## Планы по развитию своего форка

- ✓ Соберем выделенную команду по работе над форком;
- ✓ Будем развивать комьюнити вокруг форка и на русском, и на английском;
- ✓ Перепишем open source документацию Greenplum в полном объеме;
- ✓ Составим и спланируем roadmap для нашего форка.



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

Новости

Планы по Greenplum 7

Планы по развитию  
своего форка

## Наш форк Greenplum

[github.com/arenadata/gpdb/](https://github.com/arenadata/gpdb/)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

Новости

Планы по Greenplum 7

Планы по развитию  
своего форка

## Ветка 6.x

[github.com/arenadata/gpdb/tree/adb-6.x](https://github.com/arenadata/gpdb/tree/adb-6.x)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

Новости

Планы по Greenplum 7

Планы по развитию  
своего форка

## Ветка 7.x

[github.com/arenadata/gpdb/tree/adb-7.1.0](https://github.com/arenadata/gpdb/tree/adb-7.1.0)



Arenadata DB

Вклад в Open Source

Enterprise Edition

Мониторинг запросов

Управление бэкапами

Новости

Планы по Greenplum 7

Планы по развитию  
своего форка

## Arenadata DB

Подробнее о продукте

[docs.arenadata.io](https://docs.arenadata.io)



[Arenadata DB](#)

[Вклад в Open Source](#)

[Enterprise Edition](#)

[Мониторинг запросов](#)

[Управление бэкапами](#)

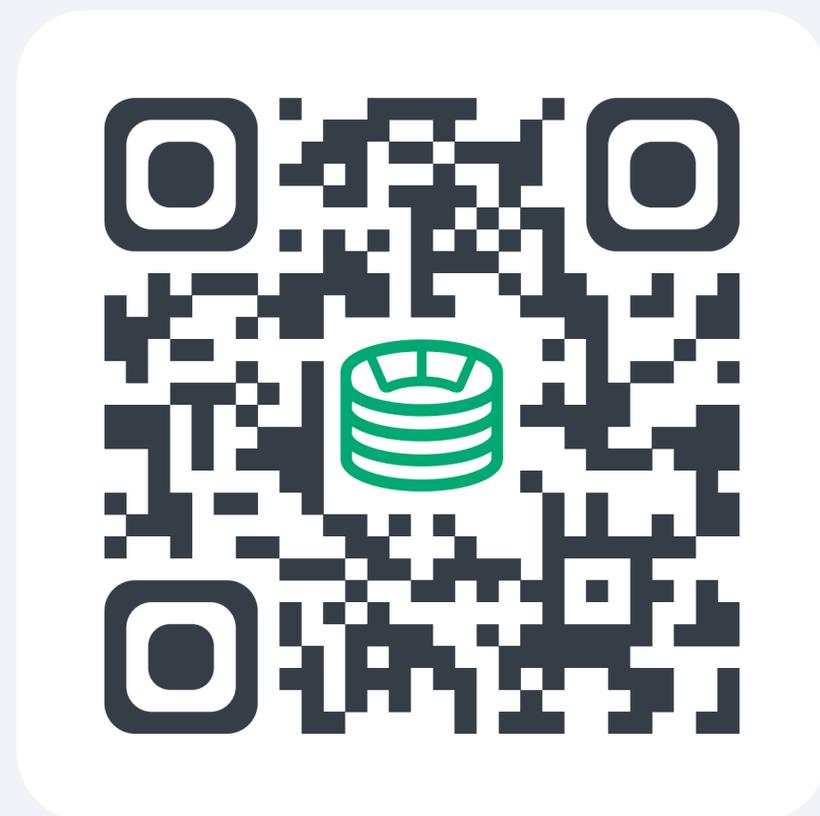
[Новости](#)

[Планы по Greenplum 7](#)

[Планы по развитию  
своего форка](#)

[Вопросы и ответы](#)

**Спасибо за внимание.  
Вопросы?**



[Arenadata DB](#)[Вклад в Open Source](#)[Enterprise Edition](#)[Мониторинг запросов](#)[Управление бэкапами](#)[Новости](#)[Планы по Greenplum 7](#)[Планы по развитию  
своего форка](#)[Вопросы и ответы](#)

## На что стоит обратить внимание в Greenplum 7

- ✓ Переделан механизм работы ресурсных групп;
- ✓ Проблема с механизмом analyze АО-таблиц;
- ✓ ORCA функционально отстает от Legacy:
  - Многоуровневое партицирование;
  - Нет поддержки новых возможностей **fdw**;
  - Нет поддержки партицирования по **hash**, по **range**;
  - RLS не поддерживается;
  - UPSERT;
- ✓ Нет поддержки логической репликации;
- ✓ Удалено расширение **gpperfmon**;
- ✓ Версия ещё не опробована на больших кластерах (для примера, стабилизация Greenplum 6 состоялась к версии 6.11);
- ✓ Проект в upstream заархивирован.

