

ЭКСПЕРТЫ В ОБЛАСТИ
АНАЛИТИЧЕСКИХ РЕШЕНИЙ



APIENS
s o l u t i o n s



ARENADATA

Миграция в Arenadata DB

Крючков Дмитрий
Исмаилов Дмитрий
Александр Тарасов

- ✓ Основана в **2012**
- ✓ Эксперты в разработке аналитических решений, хранилищ данных, систем планирования и бюджетирования
- ✓ **100+** инженеров данных
- ✓ **30+** заказчиков (М.видео, Детский мир, Комус, Лента, ВТБ, Сбербанк, Альфа-банк и др.)
- ✓ Технологические платформы Arenadata, SAP, Superset, Airflow



Arenadata

Российский вендор платформы больших данных на базе технологий Open Source



СТАБИЛЬНОСТЬ

- **5 лет** на рынке;
- **220** человек;
- **2 млрд Р** выручка 2022.



50+ ЗАКАЗЧИКОВ

- X5, Магнит, Ашан, Ривгош, Комус, Burger King;
- ВТБ, ПСБ, ДОМРФ, Росбанк, ОКБ, Синара, Ак Барс банк;
- ФНС, ДИТ МСК, ПФР, НСУД;
- ГПН, ММК, Норникель.



МАСШТАБ

- **30+ Петабайт** данных на **60.000+ ядер** работает на Arenadata у **50+** заказчиков в РФ и СНГ.



ON-PREM & CLOUD

- Onpremise и ПАКи;
- PaaS: VK Cloud;
- IaaS: MTC Cloud, Sber Cloud;
- Гетерогенные ландшафты;
- Гособлако: ГЕОП+.



ФОКУС БИЗНЕСА

- **100%** бизнеса Arenadata – создание СУБД на базе Open Source;
- Поддержка 1, 2 и 3 линии;
- Архитектурный надзор от вендора.



ЭКСПЕРТЫ

- **30+** технологических партнеров и интеграторов;
- **1.200+** обученных специалистов и экспертов.



COMMUNITY

- **2 место в мире** среди контрибьютеров в Greenplum;
- **700+** фич в Hadoop в год;
- Профессиональная [документация](#).



РЕГУЛЯТОРЫ

- **Отечественные ОС**;
- **Реестр** Минцифры;
- **Требования** регуляторов.

Место Arenadata в архитектуре





Полная миграция

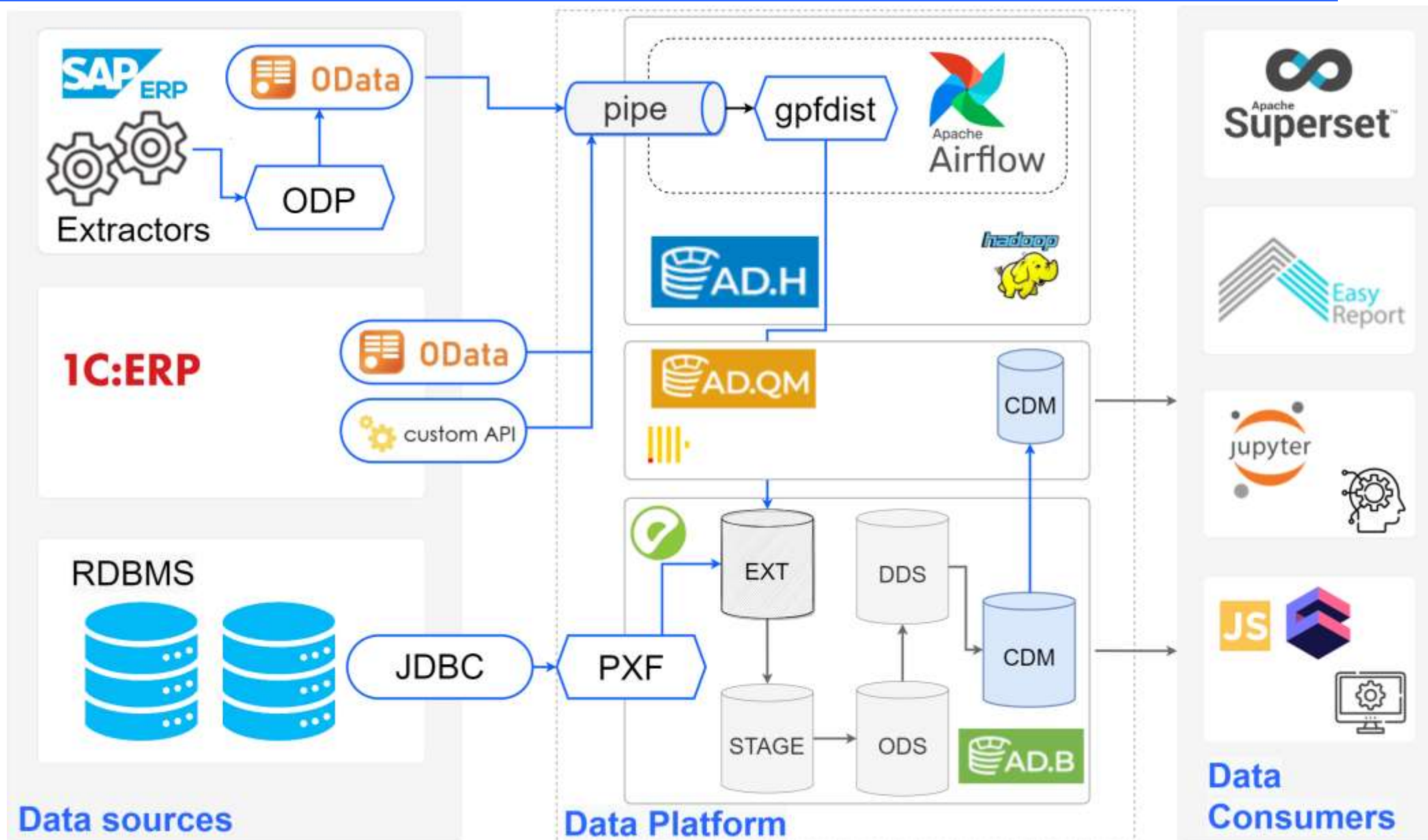
Замена всей платформы данных и SAP BI



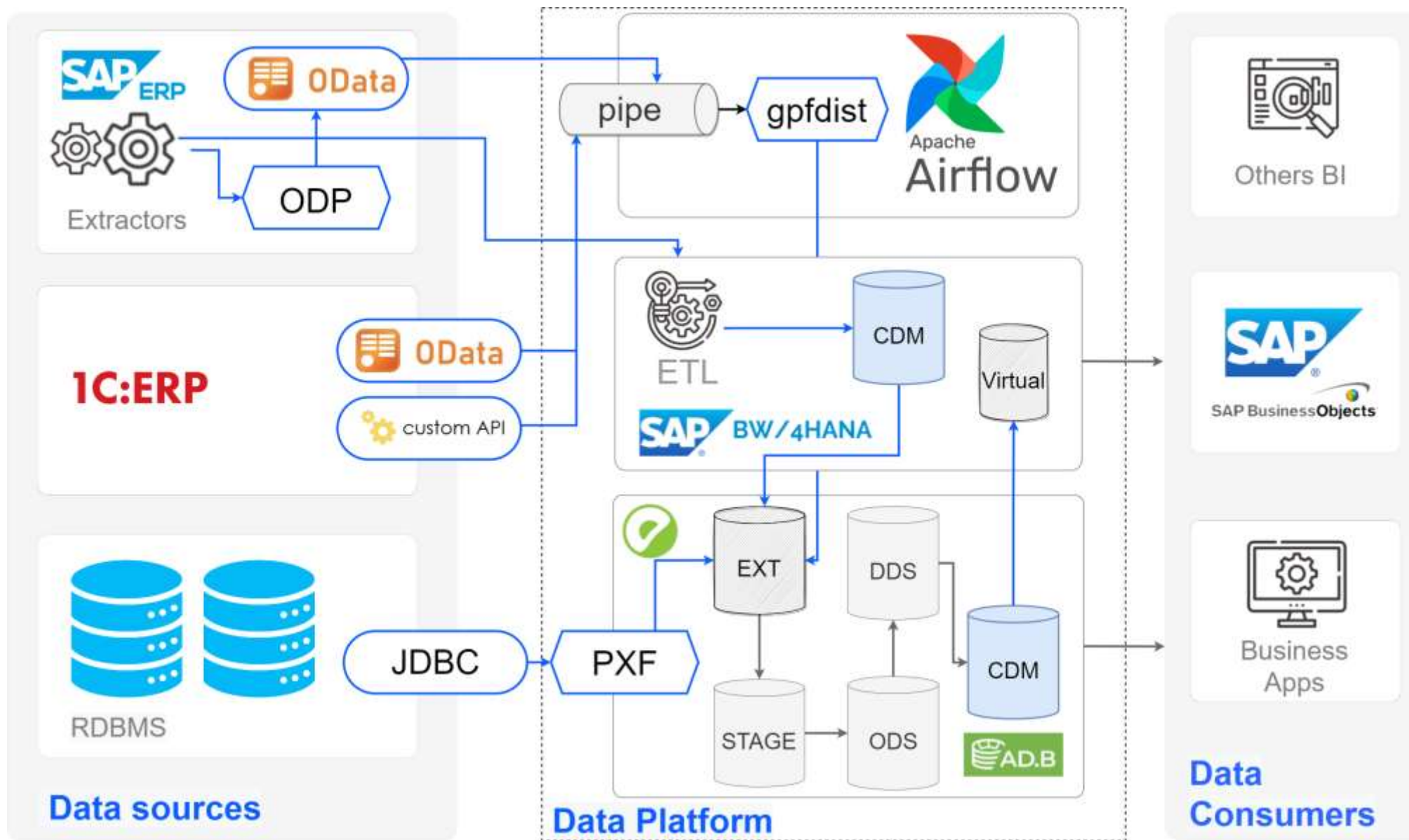
Гибридная архитектура

Максимальное переиспользование и плавная трансформация

Целевая архитектура: Arenadata Platform



Гибридная архитектура



Последовательная миграция функционала



Миграция модели данных BW - ADB



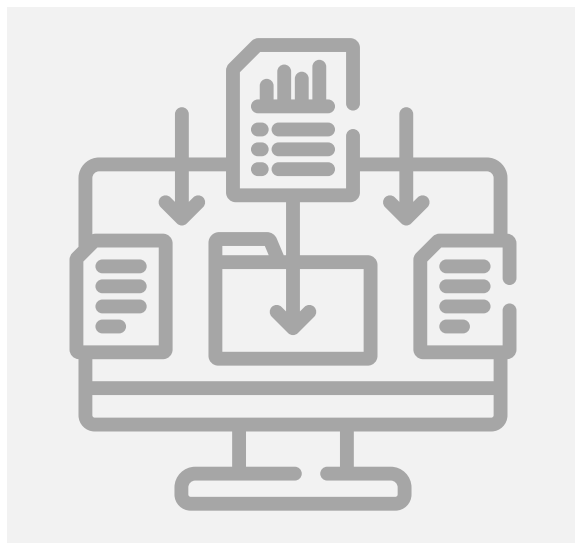
Выгрузка данных BW-ADB



Настройка инкрементального обновления из SAP систем



Настройка ETL/ELT для ADB



**Настройка источников
OData**



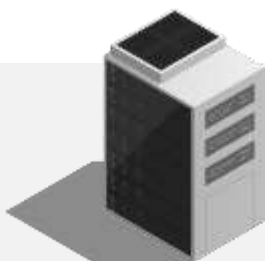
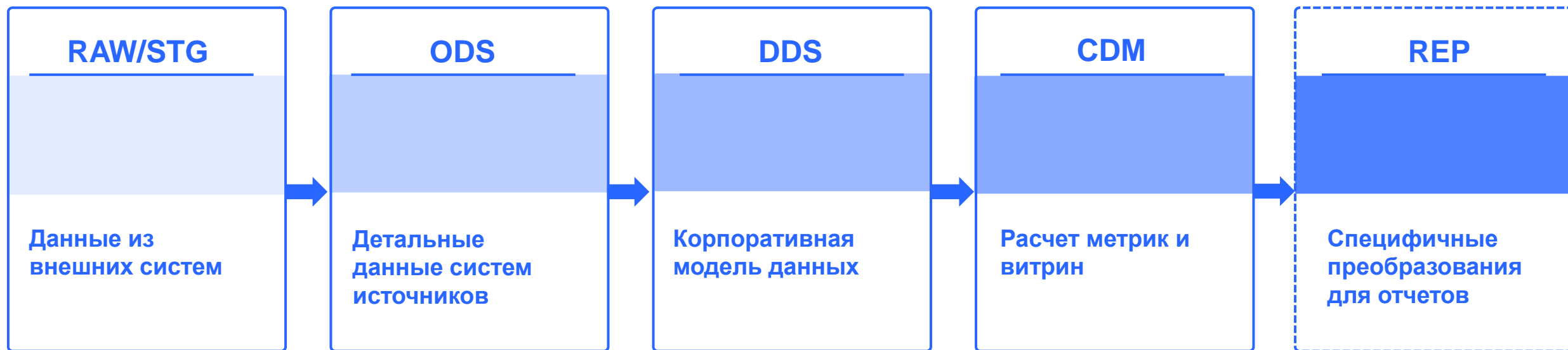
**Реализация загрузки
на Airflow, используя
gpfdist и pipe**



**Подходы к загрузке в
ADB и ETL фреймворк**



Слои данных



Модели данных:

Денормализованная, 1НФ, Звезда, 3НФ, Снежинка, DataVault, Anchor Modelling



Транзакция SEGW

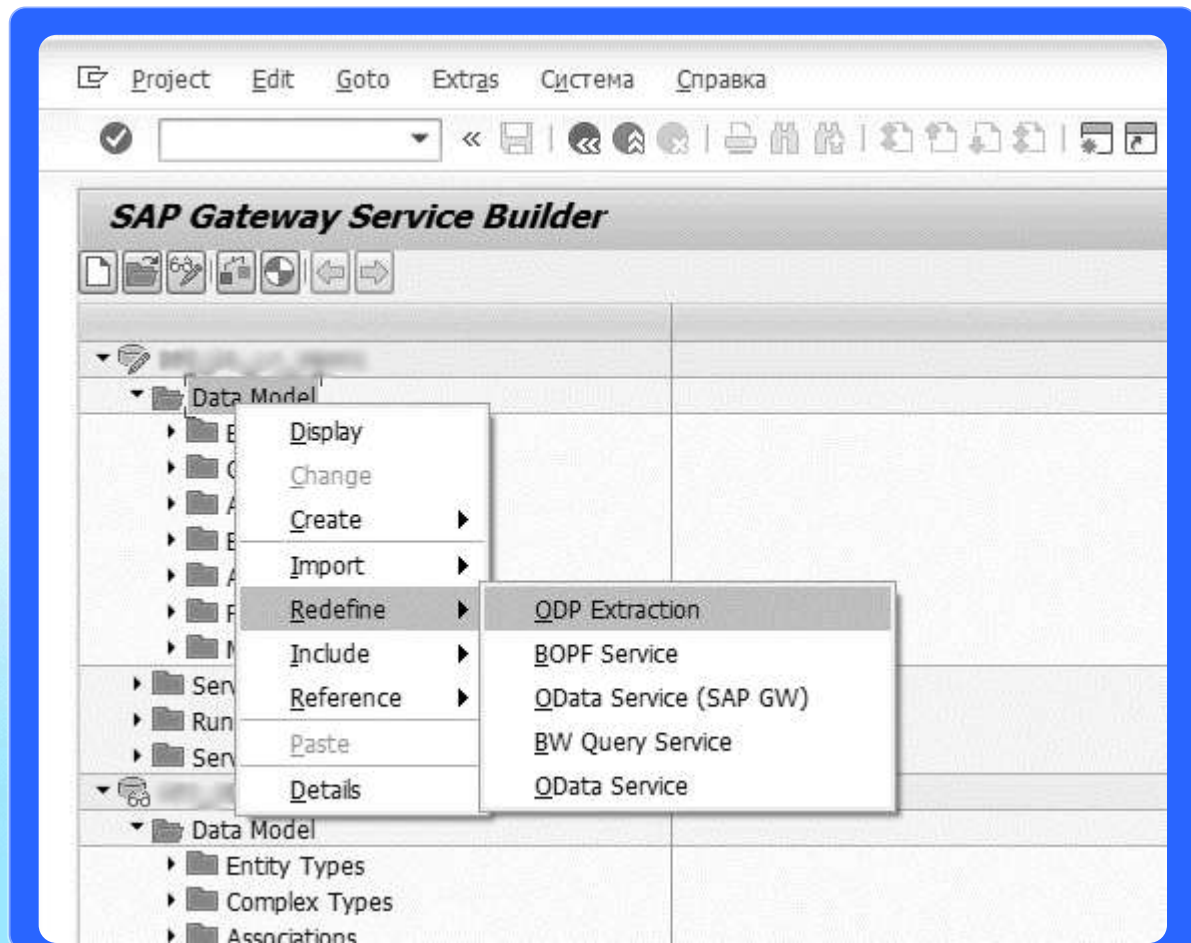
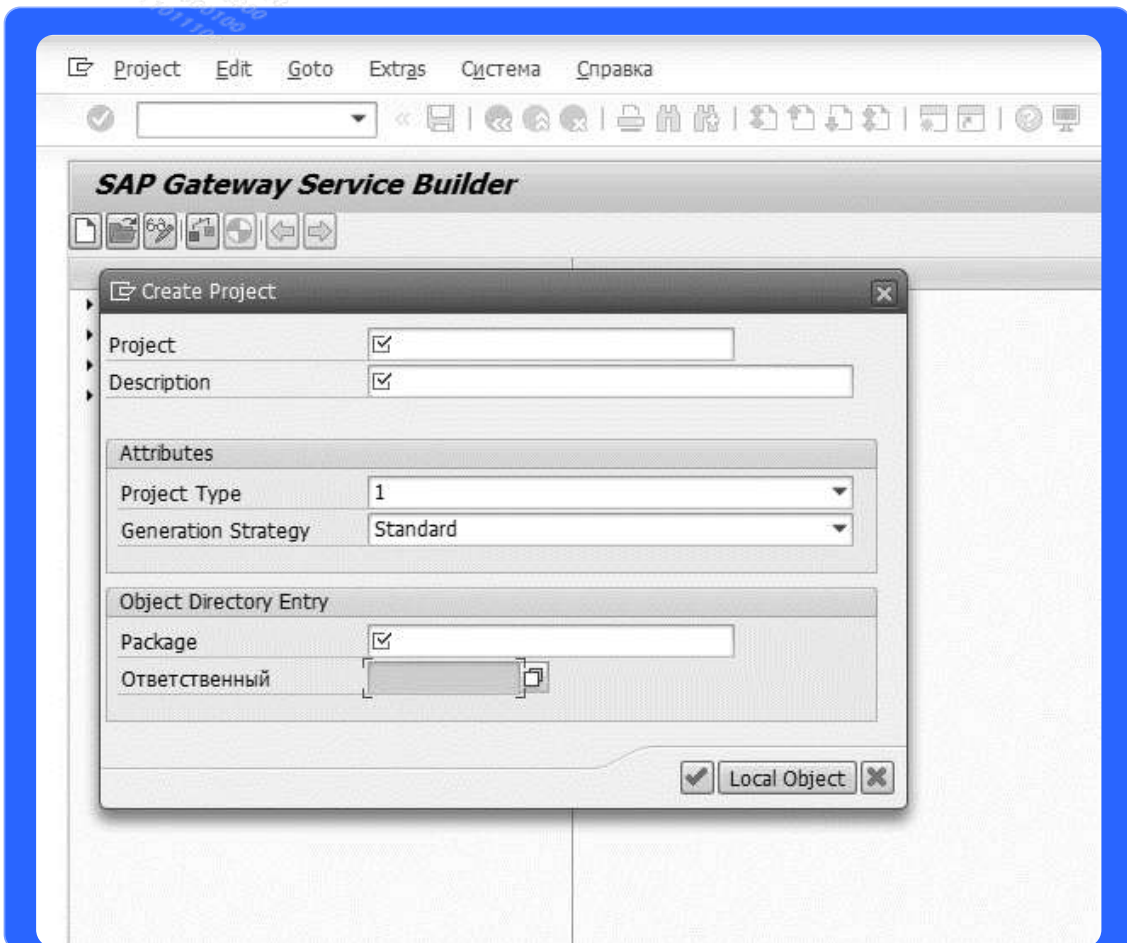
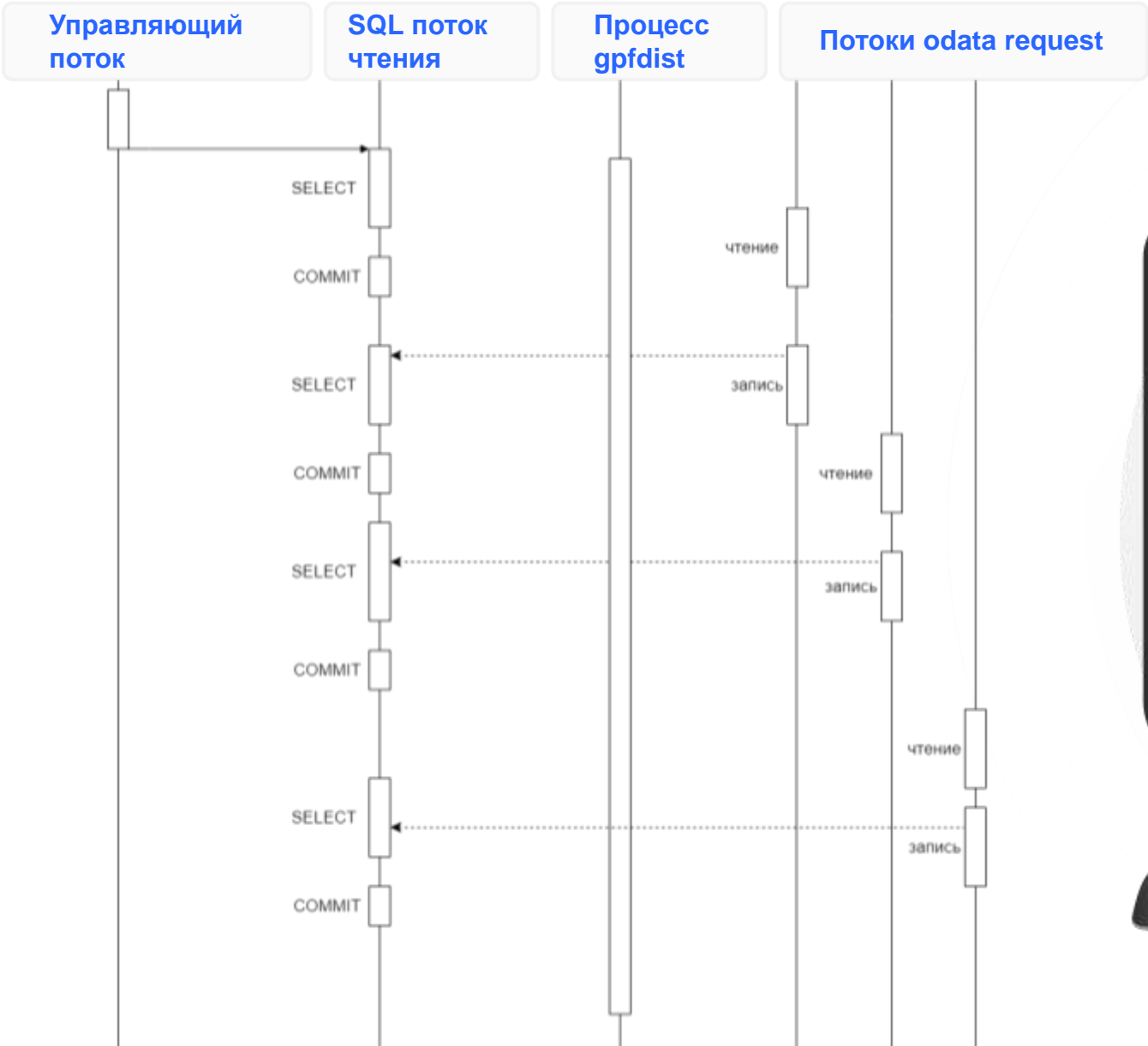


Схема потоков экстракции



- **Ограниченность в применении привычных операций**

Только добавление записей в большие таблицы без обновлений по ключу

- **Отсутствие оптимизации ETL-инструментов для ADB**

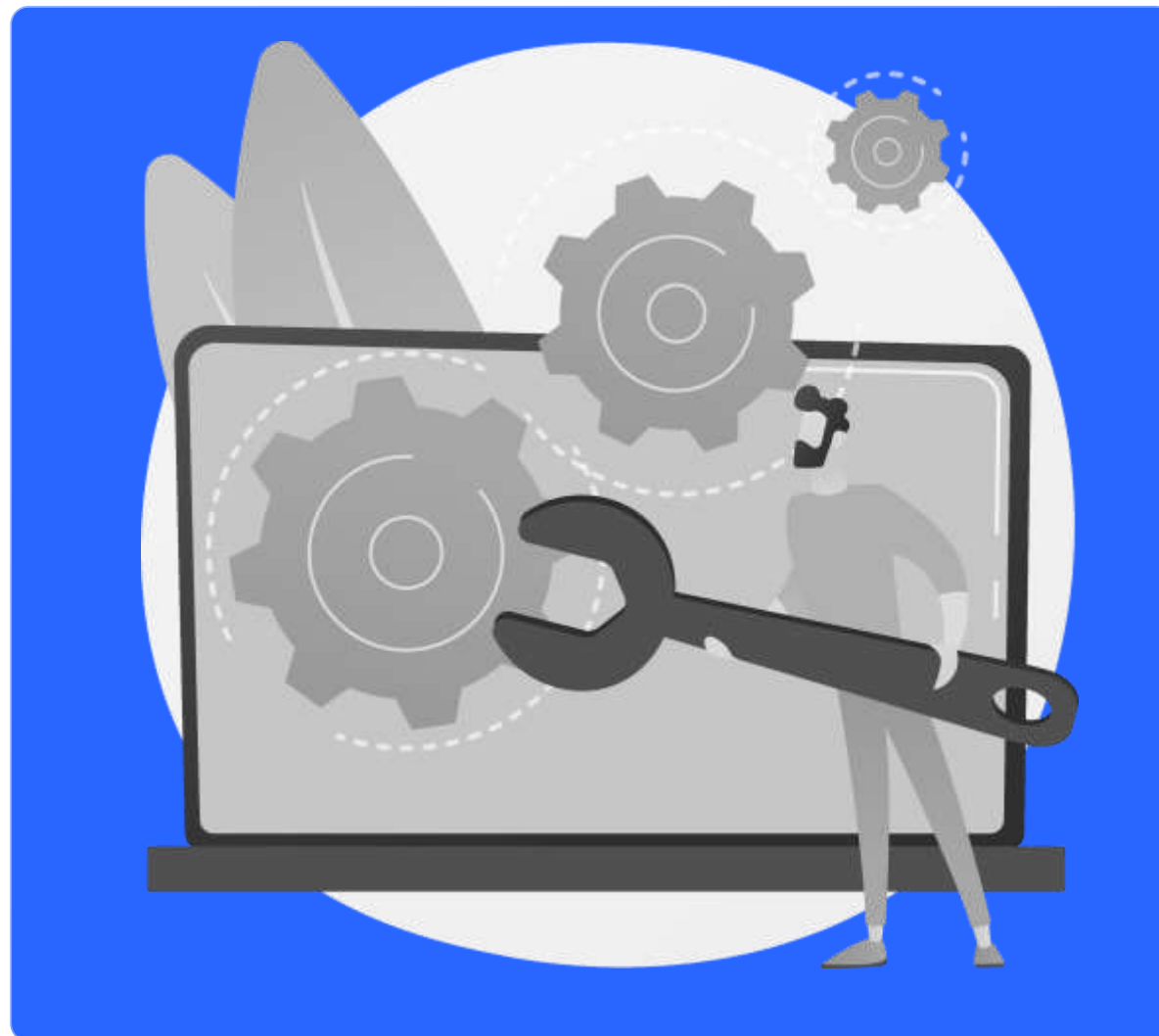
Необходимо использовать rxf или gpfdist

- **Отсутствие инструментов мониторинга и логирования**

Изначально доступна только sql-консоль

- **Долгое внедрение собственных разработок**

При разработке с нуля





- **Быстрое развертывание**
и первые результаты
- **Минимальное количество настроек**
- **Простота разработок**
и прозрачный код plpgsql + python
- **Оптимальная загрузка**
с учетом особенностей ADB
- **Рост компетенций**
в соответствии с объемом миграции
- **Оркестрация**
Логирование и мониторинг

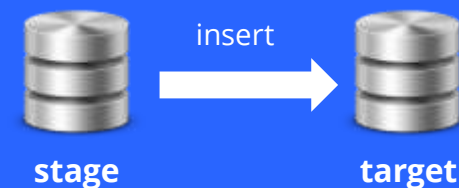
Способы загрузки

- Полная перезапись
- Чистая дельта



Полная перезапись

- ✓ Формирование таблицы с бэкапом данных целевой таблицы
- ✓ Очистка целевой таблицы
- ✓ Вставка данных в целевую таблицу



Чистая дельта

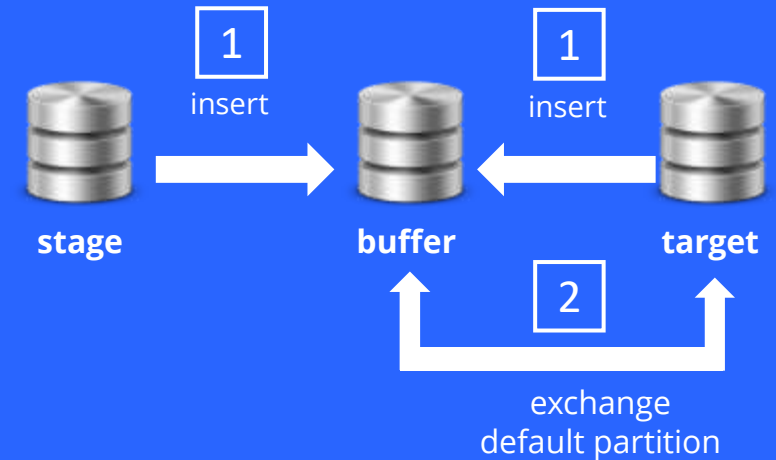
- ✓ Вставка дельты данных в целевую таблицу

Способы загрузки

- Изменения по ключу



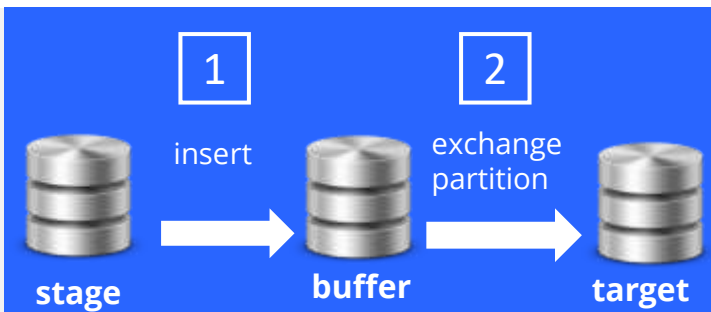
- ✓ Удаление данных из целевой таблицы по ключу на основании новых данных
- ✓ Вставка дельты данных в целевую таблицу



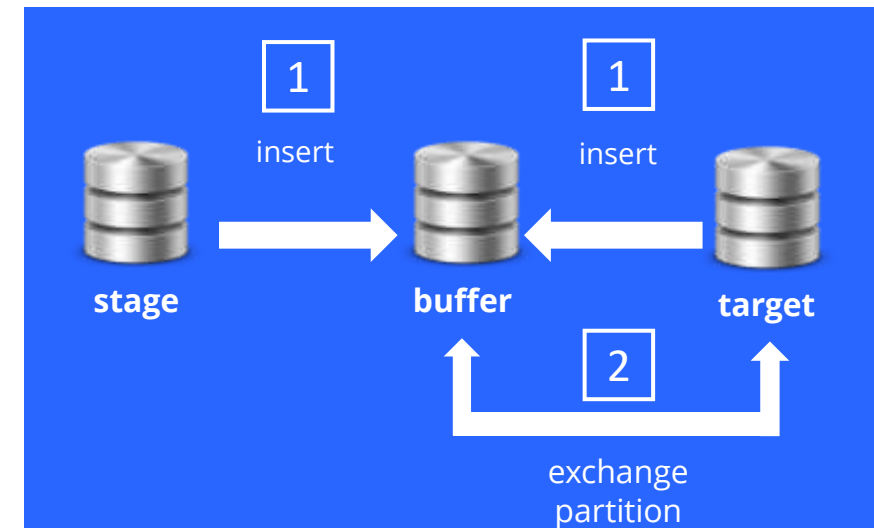
- ✓ Вставка текущих данных из целевой таблицы и новых данных дельты в буферную таблицу со сравнением по ключу
- ✓ Замена партиции по умолчанию в целевой таблице данными буферной таблицы

Способы загрузки

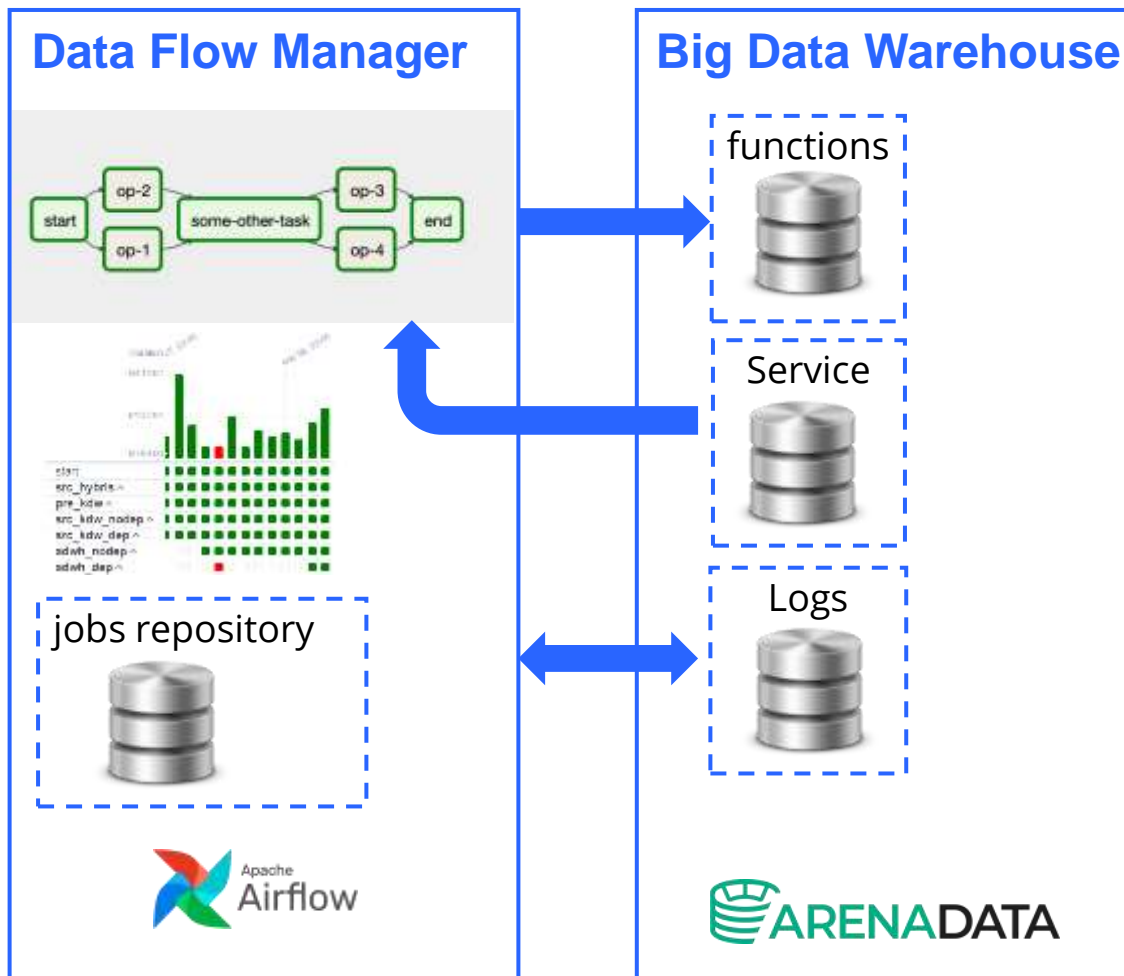
- Полный интервал бизнес дат
- Дельта за интервал бизнес дат



- ✓ Формирование буферной таблицы с новыми данными одной партиции
- ✓ Замена нужной партиции в целевой таблице



- ✓ Вставка текущих данных из целевой таблицы и новых данных в буферную таблицу за интервал одной партиции со сравнением по ключу
- ✓ Замена нужной партиции в целевой таблице

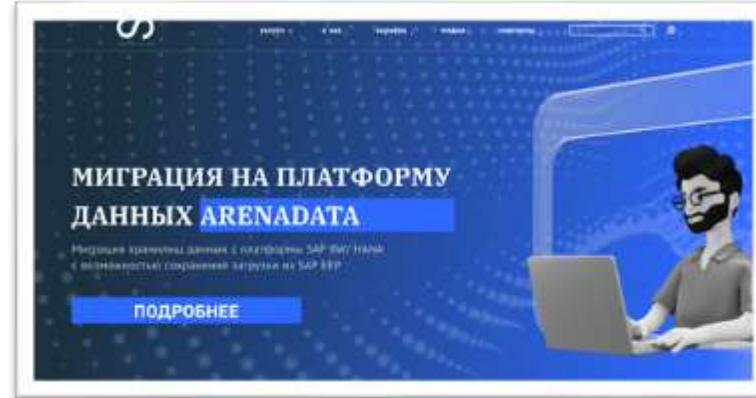


- Автоматическая генерация блоков загрузки данных на основании настроек
- Логирование в оркестраторе и базе
- Оперативное отслеживание статуса загрузок

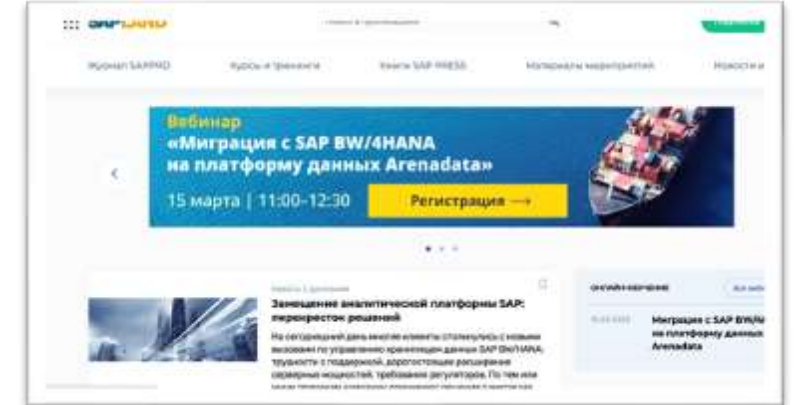
Q & A



[Как с помощью импортозамещающих технологий снизить нагрузку на SAP HANA](#)



[Sapiens Solutions](#)



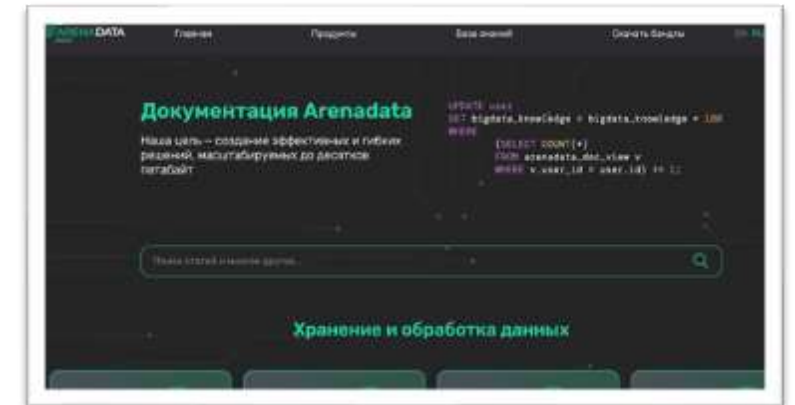
[SAPLAND](#)



[YouTube Arenadata](#)



[Telegram Arenadata](#)



[Возможности решений Arenadata](#)

Контакты

- <https://sapiens.solutions>
info@sapiens.solutions
+7495 215 1757
- Юрий Ефаров:
y.efarov@sapiens.solutions
<https://t.me/yefar>
- Константин Ракитин:
c.rakitin@sapiens.solutions
<https://t.me/keizercos>
- <https://arenadata.tech/>
info@arenadata.io
- sapland@sapland.ru
<https://sapland.ru/>