



# КАК МИГРИРОВАТЬ

---

100500 пользовательских  
объектов из старого  
хранилища в новое и  
остаться в ЖИВЫХ

# Sandbox

## Песочница

Набор ресурсов, которые позволяют специалистам экспериментировать и изменять данные любым способом.

Выделенная часть корпоративного хранилища данных или витрины данных.

Возможность напрямую объединять производственные данные и данные самой песочницы.

Используется ограниченным кругом пользователей. Содержит данные, отделённые от производственного хранилища. Количество объектов может достигать 1000+



# О ЧЁМ ПОГОВОРИМ

## 1 Планирование

Начните заранее- анализа будет много.

1

## 3 Автоматизируй это

Подходы к переносу данных в новое хранилище.

3

## 5 Про людей

Новое хранилище с человеческим лицом

5

## 2 Подход «мягкой силы»

Как не напугать пользователей на старте проекта.

2

## 4 Договариваемся на берегу.

На какие грабли можно (не)наступить.

4

# Планирование и анализ



## Анализ параллельно выбору целевого хранилища

На этом этапе выбор финального целевого хранилища может быть не завершён



## Определить перечень песочниц

Выявить недовыявленное



## Выявить владельцев и пользователей

В случае отсутствия владельца- интервью с пользователями для выявления нового владельца



## Формализовать результаты (BPM/ServiceDesk/C3)

Документировать проведённый анализ (task tracker, wiki, Confluence)

# Планирование и анализ

✓ Выгрузка мета-данных из дата-словаря  
текущего хранилища

✓ В полученной выгрузке актуализировать и  
валидировать с пользователями:

Владельцев песочниц

Список объектов (удалить неактуальные и не  
тратить ресурсы на перенос)

Глубину хранения данных в объектах (уменьшить  
излишнюю глубину и сократить ресурсы на  
перенос и хранение в новом хранилище)

Порядок переноса объектов (выявить кандидатов  
на приоритетный перенос)

Процессы, построенные на объектах (собрать  
скрипты)

✓ Выявить команды аналитиков, у которых в  
песочницах реализованы сложные  
нерегламентированные процессы. Возможно,  
понадобится помощь в миграции кода на  
новое хранилище.

1	cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster	some-cluster
2	business_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
3	db_name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name	some-db-name
4	tab_name	some-table-name	some-table-name2	some-table-name3	some-table-name4	some-table-name5	some-table-name6	some-table-name7	some-table-name8	some-table-name9	some-table-name10
5	creator_name	some-user	some-user	some-user	some-user	some-user	some-user	some-user	some-user	some-user	some-user
6	create_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
7	byte	1548288	10788352	8978432	479713280	10089472	1588736	2192896	518144	518144	123988992
8	max_select_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
9	min_select_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
10	max_update_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
11	min_update_dt	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY	DD.MM.YYYY
12	total_cnt	8024	1265016	4370	115	1	806	2656871	25616	1809	354
13	user_total_list	{"some-user"}	{"some-user", "some-user2"}	{"some-user"}	{"some-user3"}	{"some-user"}	{"some-user"}	{"some-user"}	{"some-user"}	{"some-user"}	{"some-user"}

# Подход к миграции



## Подход «мягкой силы»

- Определена целевая дата переноса.
- Данные переносятся поэтапно.
- Старое хранилище продолжает работать в течение переходного периода.
- Новые объекты создаются только на новом хранилище



## Приоритеты

- Совместно с пользователями-выбрать список приоритетных объектов
- Миграция данных и миграция кода



## Code Freeze

- Будет ли применён для пользовательских песочниц
- Эксплуатируется ли старое хранилище до завершения переноса данных



## Новые объекты

- С момента проведения анализа в песочницах появились новые объекты
- Перед стартом процесса переноса-повторная выгрузка объектов старого хранилища
- Провалидировать с владельцем список переносимых объектов



## Архивные периоды

- Для больших объектов- согласовать с владельцами перенос данных за архивные периоды (если текущие процессы от них не зависят)
- Так пользователи смогут протестировать функционал нового хранилища на привычных данных



## Задачи сверки

- Валидировать с владельцами песочниц механизмы сверки данных в хранилищах
- Автоматический компаратор объектов в разных источниках

\*Задача «со звёздочкой»: аудит наличия в песочницах персональных данных (шире-информации ограниченного доступа)

# Автоматизация переноса данных

## ✓ Подготовительная работа

Совместно с архитекторами определить, какие опции нового хранилища будут активно применяться (колоночная модель хранения, разные варианты сжатия, ключи распределения и т.д.).

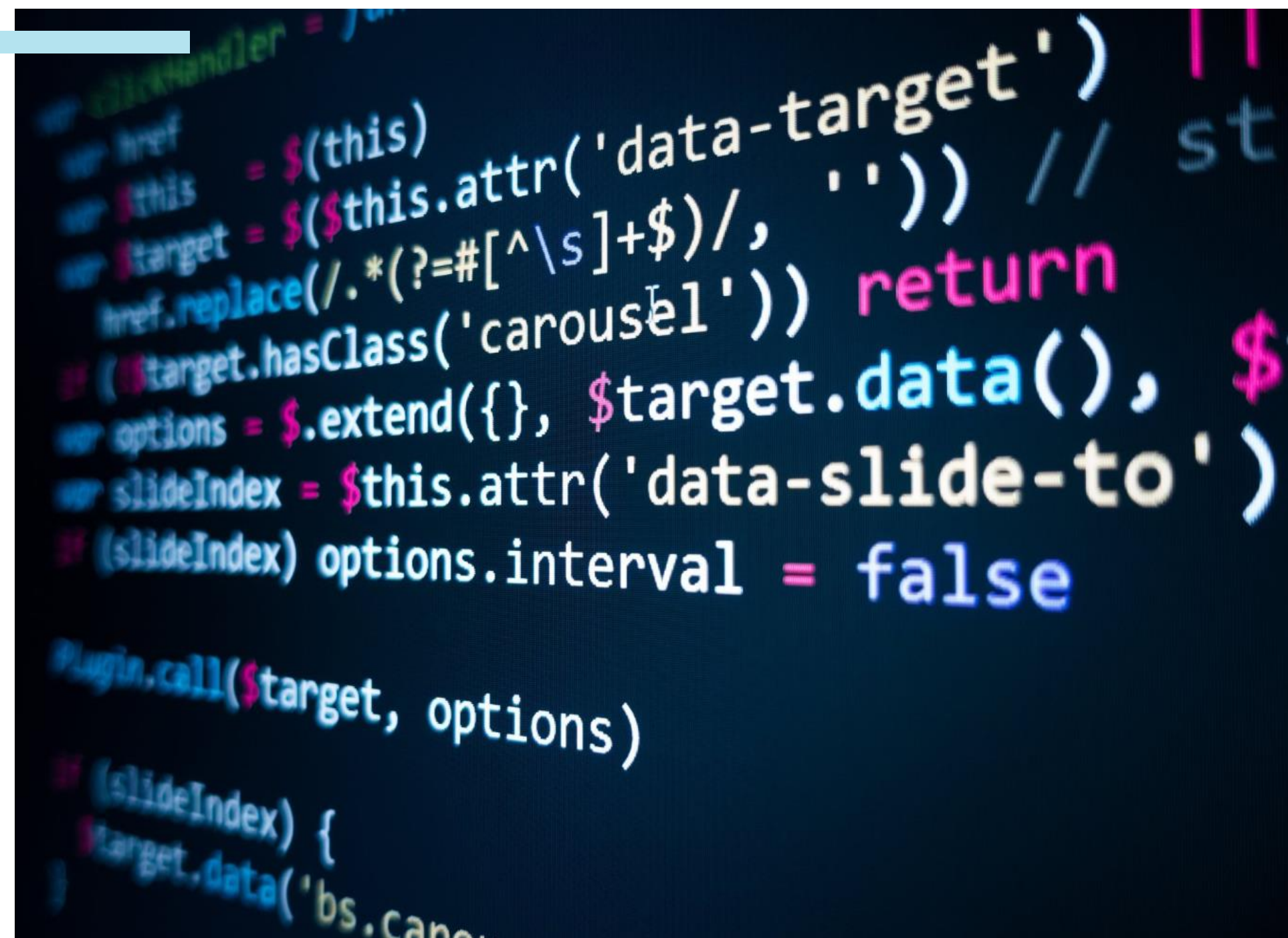
Составить маппинг типов данных между новым и старым хранилищами.

Правила именования объектов после миграции (case sensitive, кириллица в наименовании полей, имена, превышающие допустимую длину в новом хранилище, наличие спец.символов в названии полей).

Определить, какую максимальную длину текстовых полей поддерживает новое хранилище.

Определить, какие механизмы экспорта данных доступны (fastexport, bulkload и т.д.)

Определить, появляется ли необходимость изменять ролевую модель доступа к данным в новом хранилище.



# Этапы автоматизации





# На какие грабли можно (не) наступать

- ✘ Выявление объектов на перенос, не обследованных на этапе анализа по причине размещения в общекорпоративной песочнице.
- ✔ Обследовать общедоступную разделяемую песочницу в компании (common, exchange, general и т.д.). Провести анализ определения принадлежности объектов конкретным подразделениям. Определить подход к переносу таких объектов.
- ✘ Ошибки при работе процесса переноса, связанные с необследованными типами данных.
- ✔ Не ограничиваться семплом данных для анализа.
- ✔ Выгрузить из словаря данных все используемые типы для последующей автоматизации.
- ✘ Сбои при линейном переносе объёмных таблиц (50Гб+) и отсутствие возможности продолжить с момента останова.
- ✔ Предусмотреть при разработке автоматизированного решения попартиционный перенос (задавать параметрами поле партиционирования и размер переносимой партии).



# Поддержка пользователей

- ✓ Публикация на общедоступном ресурсе компании информации о проводимых работах по песочницам. Ответственные и статус.
- ✓ Дополнительно размещение страницу «Руководство пользователя...» для нового хранилища. Кроме документации, добавить инструкцию по подключению к новому хранилищу и реестр лучших практик по работе с ним.
- ✓ Выделение сотрудников для «бебиситтинга» пользователей. Проведение встреч с командами для обучения аналитиков подходам работы с новым хранилищем. Адаптация кодов существующих скриптов для нового хранилища. Ревью кода аналитиков/ разработчиков, консультирование.



# Спасибо!

**Гудимова Марина**

**Ведущий аналитик**

**+7 (911)350-01-58**

[m.gudimova@concept-software.ru](mailto:m.gudimova@concept-software.ru)

**Опаричев Александр**

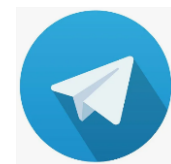
**Коммерческий директор**

**+7 (916)220-20-69**

[a.oparichev@concept-software.ru](mailto:a.oparichev@concept-software.ru)

**ООО «Концепт разработка»**

[concept-software.ru](http://concept-software.ru)



**Концепт разработка**

