

# ОБЕЗЛИЧИВАНИЕ ДАННЫХ

Спикер:  
**Алексей Даньшин**

# ПЛАН ПРЕЗЕНТАЦИИ

1. Предыстория
2. Проблемы
3. Как всё устроено
4. Существующие решения
5. Свой проект

# ПРЕДЫСТОРИЯ

## Задачи проекта

- Анализ источников витрин
- Адаптация кода под новую систему
- Получение данных для тестов
- Написание документации
- Разработка алгоритма миграции



# ПРЕДЫСТОРИЯ

## Данные для тестов

- Составить заявку на обезличивание ○
- Дождаться очереди ○
- Процесс обезличивания ○
- Согласование с ИБ ○
- Выгрузка в контур ○

# ПРЕДЫСТОРИЯ

## Проблемы обезличивания

- Сроки реализации



# ПРЕДЫСТОРИЯ

## Генератор данных

### Плюсы:

- Сохраняет связи
- Сохраняет исходный формат
- Нет ограничений по кол-ву
- Есть интерфейс
- Инструмент универсальный

### Минусы

- Не подходит для сложных витрин
- Требуется дополнительных ресурсов

# ПРЕДЫСТОРИЯ

## Проблемы обезличивания

- Сроки реализации
- Консистентность данных

# ПРЕДЫСТОРИЯ

## Проблемы обезличивания

- Сроки реализации
- Консистентность данных
- Ограничения по количеству



# ПРЕДЫСТОРИЯ

## Проблемы обезличивания

- Сроки реализации
- Консистентность данных
- Ограничения по количеству
- Потеря смысла данных и их формата



# ПРЕДЫСТОРИЯ

## Проект

Проект поддержки бизнес-идей сотрудников



Обучение Product Manager



Нашёл команду



Работа над проектом



...



Светлое будущее



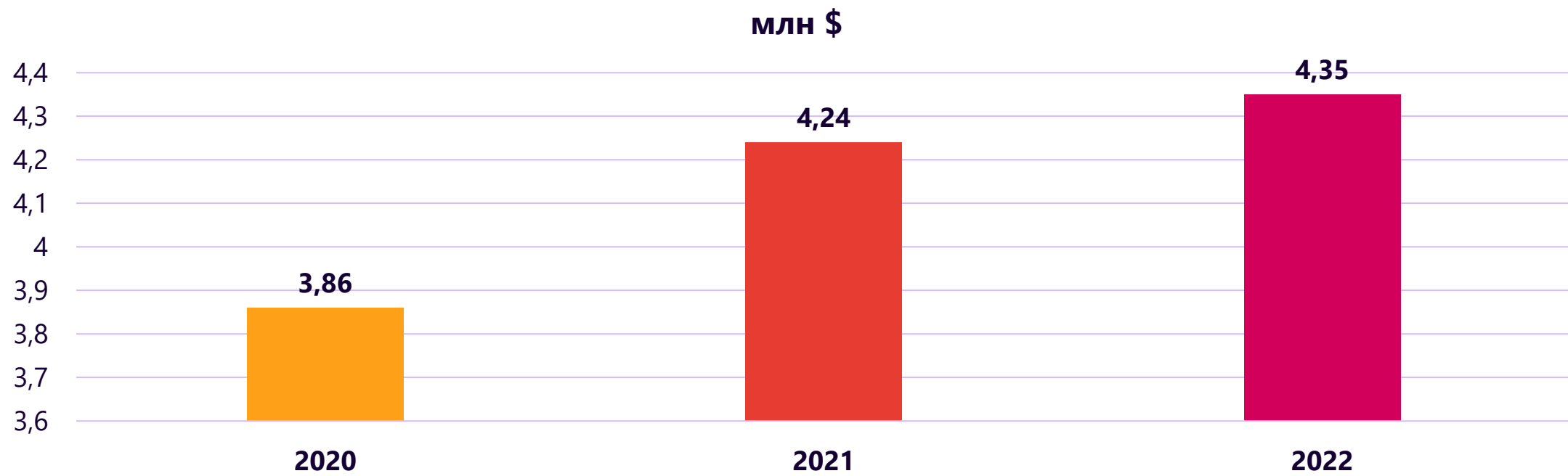
# ПРОБЛЕМА

Сохранить  
ценность данных



# СРЕДНЯЯ СТОИМОСТЬ УТЕЧКИ ДАННЫХ

Год	2020	2021	2022
млн \$	3,86	4,24	4,35



# ЦЕЛЬ

Защитить данные  
от утечки и сохранить  
их полезность.



# СПОСОБЫ РЕШЕНИЯ ПРОБЛЕМЫ

## Генерация

- Среда разработки (тестовых сред)
- Обучение пользователей
- Отладка ПО

## Обезличивание

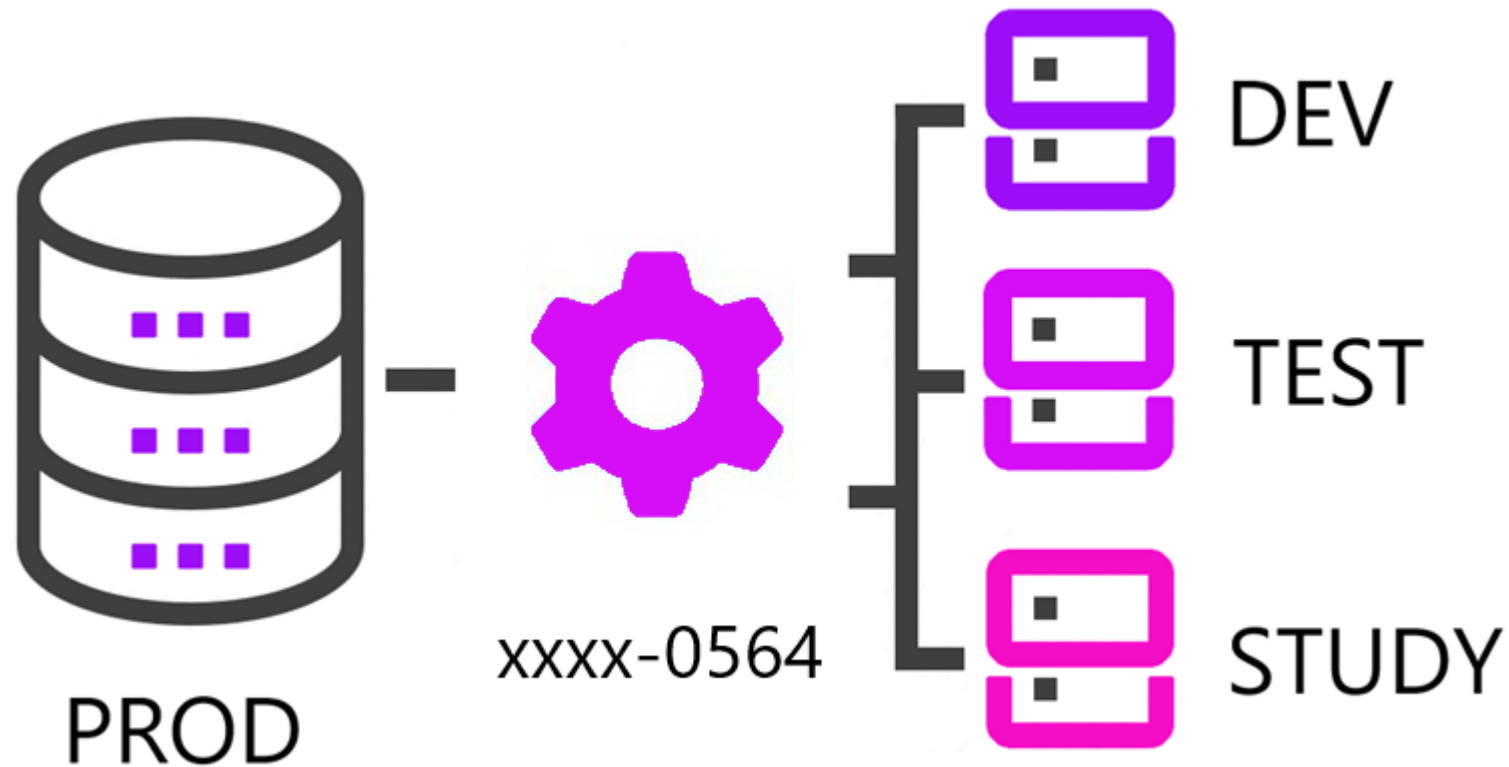
- Среда разработки (тестовых сред)
- Анализ данных
- Формирование отчётов

# ТИПЫ ОБЕЗЛИЧИВАНИЯ ДАННЫХ

- Динамический
- Статический



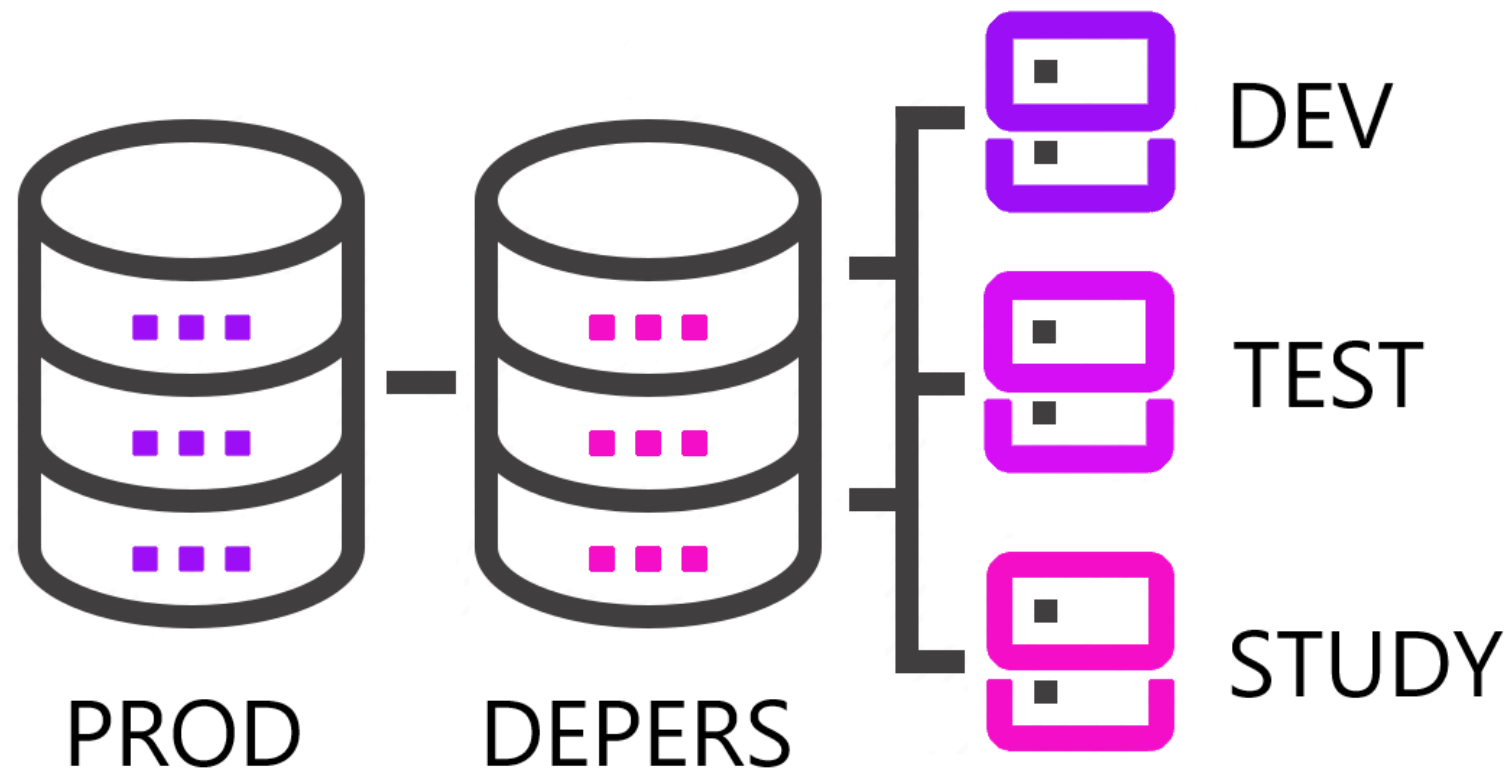
# ДИНАМИЧЕСКОЕ ОБЕЗЛИЧИВАНИЕ





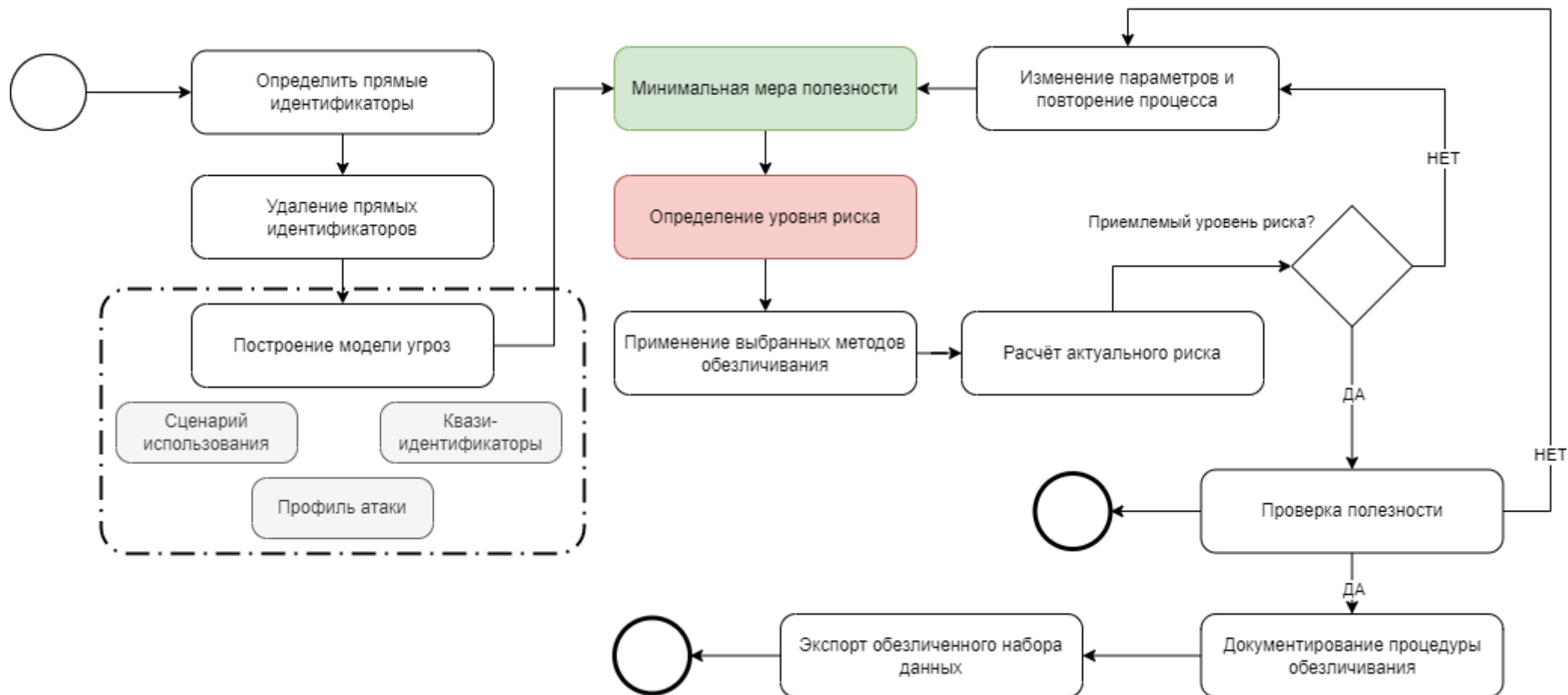


# СТАТИЧЕСКОЕ ОБЕЗЛИЧИВАНИЕ





# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ



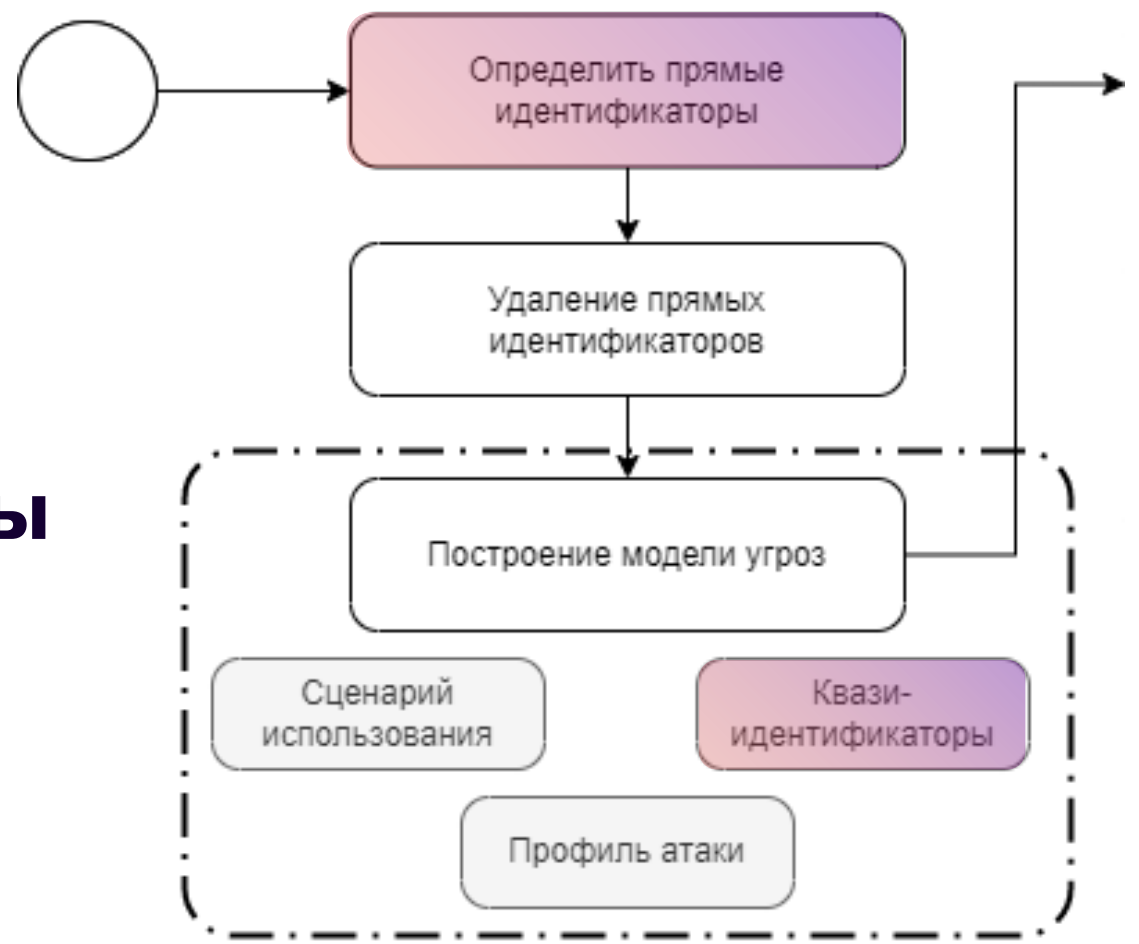
© Источник: <https://habr.com/ru/companies/rubda/articles/688116/>  
(Ассоциация больших данных)



# ПРОФИЛИРОВАНИЕ

Профилмирование данных — процесс идентификации конфиденциальных данных

- Прямые идентификаторы
- Квази-идентификаторы



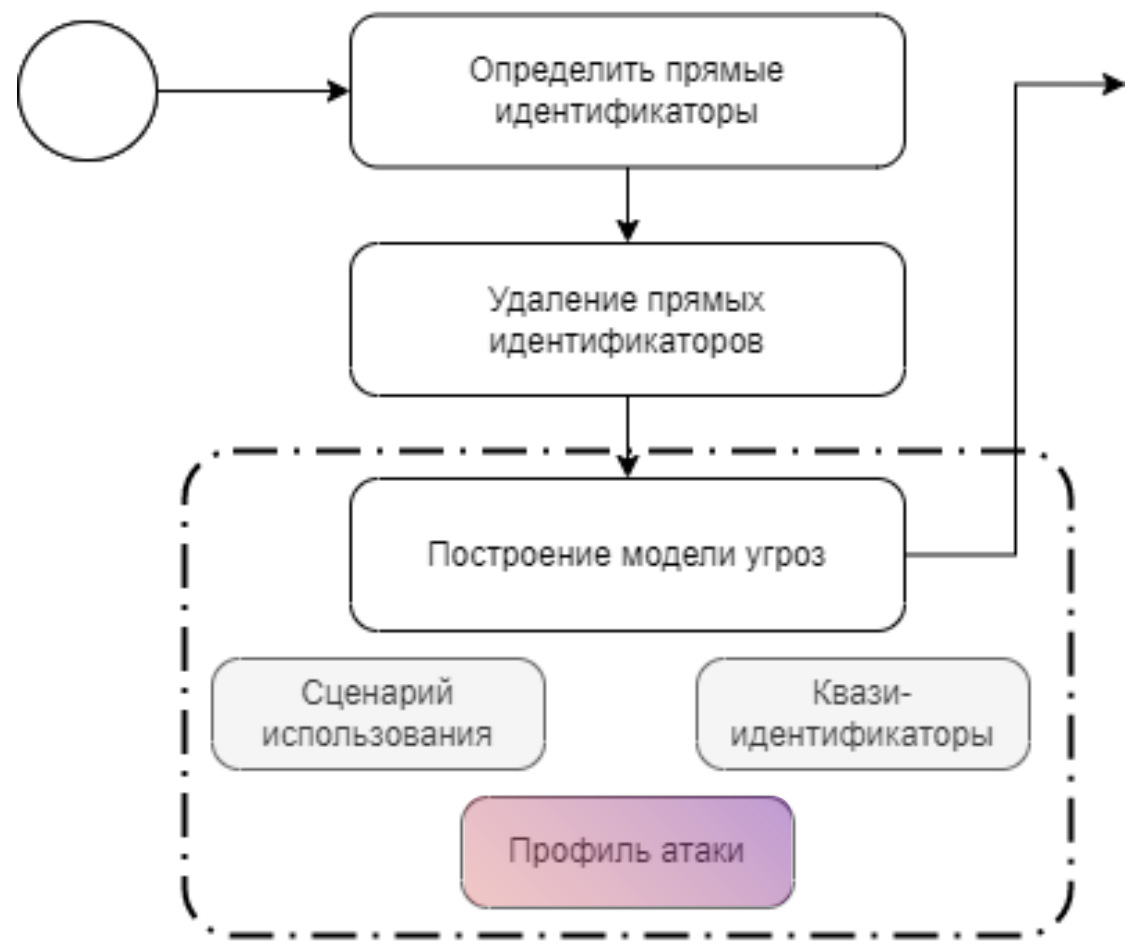


# ПРОФИЛИ АТАК

**Прокурор** – наличие информации о конкретном физ. лица.

**Журналист** – наличие общей информации.

**Маркетолог** – выделение групп лиц.





# СЦЕНАРИИ ИСПОЛЬЗОВАНИЯ

## Публичные данные

– доступ к данным не ограничен.

## Межорганизационное взаимодействие

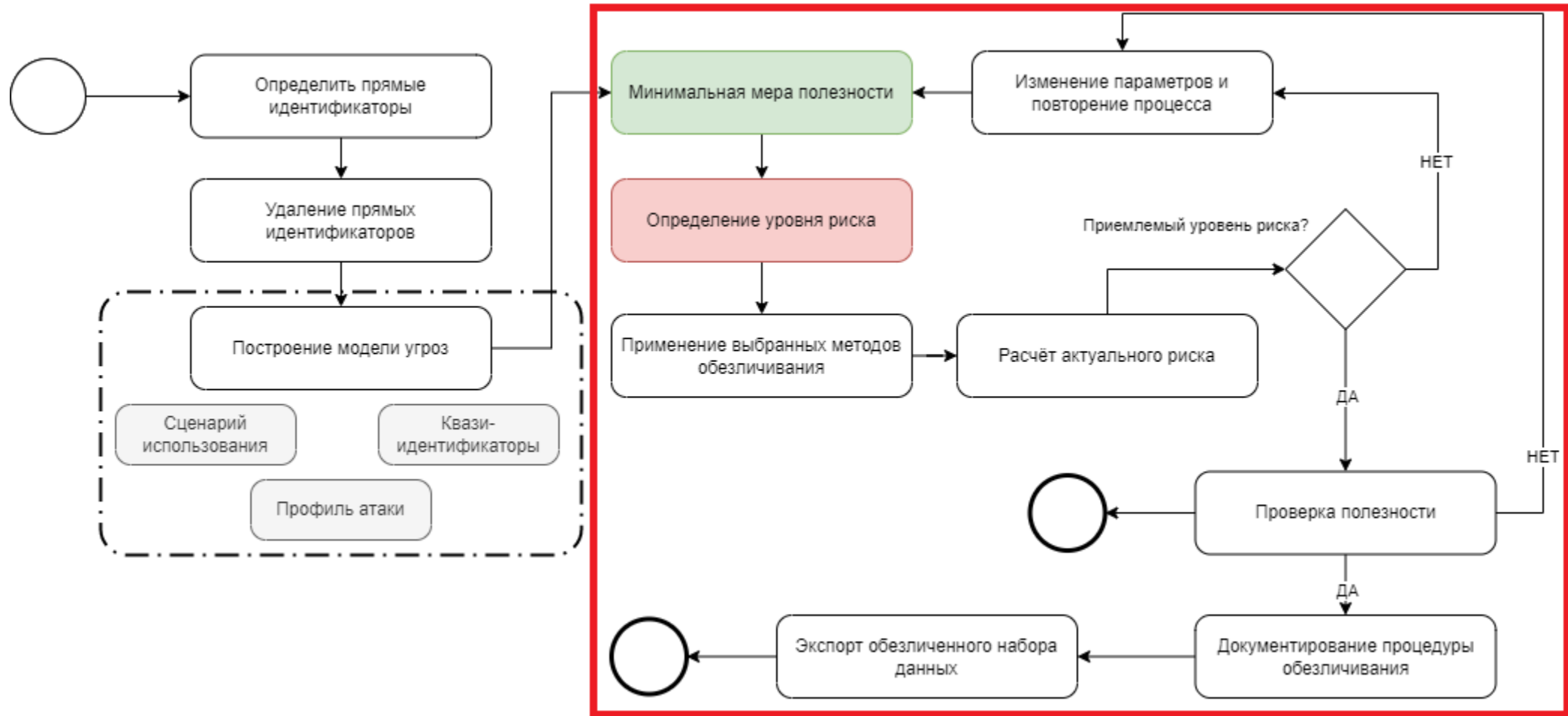
– доступ предоставляется определенному кругу лиц/организаций.

## Частные данные

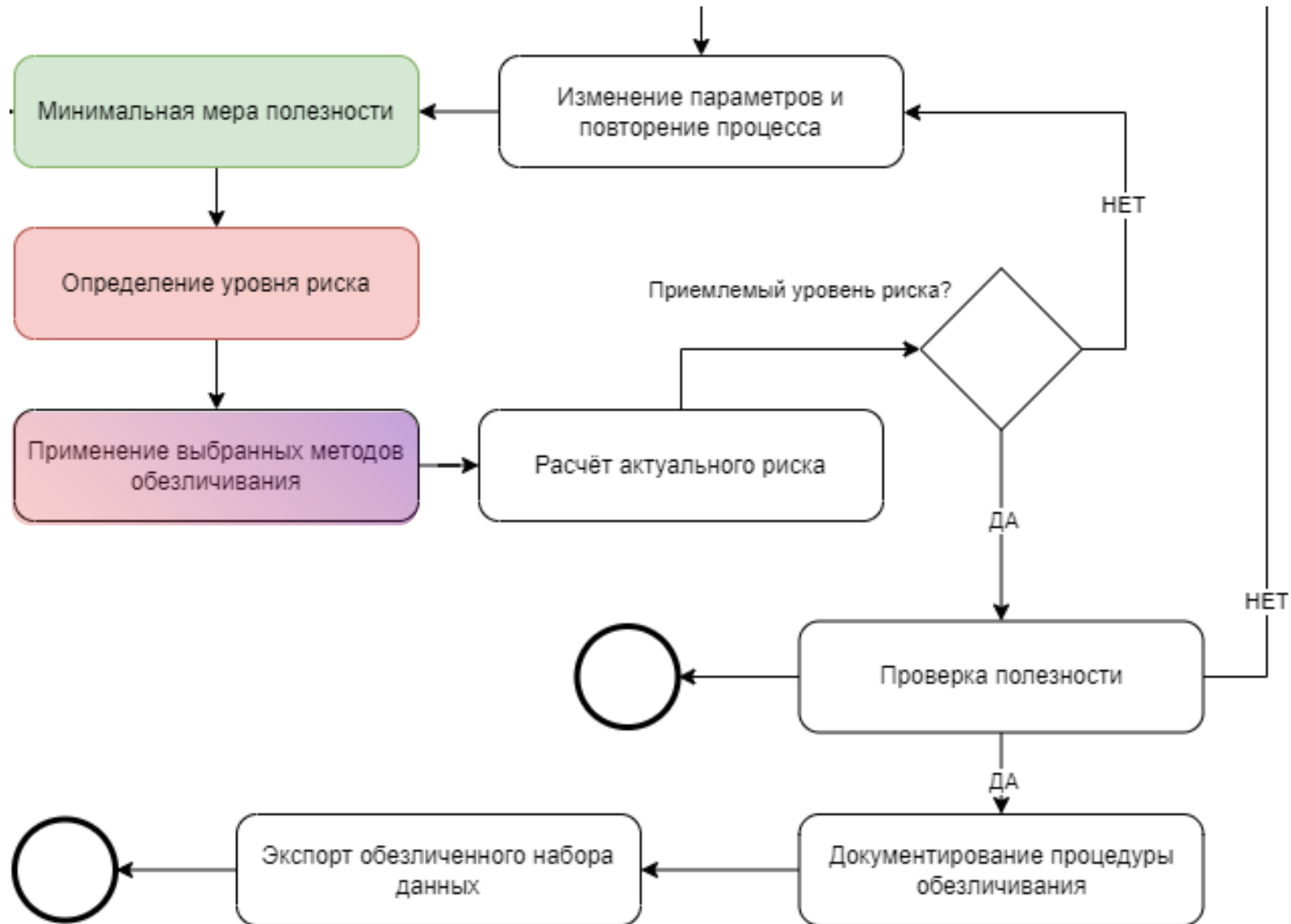
– доступ предоставляется узкому кругу людей



# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ



# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ





# Криптографические методы обезличивания

- ✓ Детерминированное шифрование
- ✓ Сохраняющее порядок шифрование
- ✓ Сохраняющее формат шифрование
- ✓ Гомоморфное шифрование





# Статистические методы обезличивания

<b>Подавление</b>	Имя	Алексей	*****
<b>Обобщение</b>	Рост	172	170-180
<b>Искажение</b>	Фамилия	Иванов	Козлов

## Перестановка

Имя	Рост	Возраст		Имя	Рост	Возраст
Алексей	172	34		*	170-180	25-35
Артём	178	28		*	170-180	25-35
Иван	175	33		*	170-180	25-35



# Статистические методы обезличивания. Подавление

## Маскирование

Имя	Рост	Серия и № пасп		Имя	Рост	Серия и № пасп
Алексей	172	4315654567		*	172	0
Артём	178	4235657454		*	178	0
Иван	175	3643789546		*	175	0



# Статистические методы обезличивания. Подавление

## Локальное подавление

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	105	Рак		*	170-180	*	Рак
Артём	178	28	Грипп		*	170-180	25-35	Грипп
Иван	220	33	ОРВИ		*	*	25-35	ОРВИ



# Статистические методы обезличивания. Подавление

## Подавление записей

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	170-180	25-35	Рак
Артём	178	28	Грипп		*	170-180	25-35	Грипп
Иван	220	105	ОРВИ		<u>*</u>	<del>170-180</del>	<del>25-35</del>	<del>ОРВИ</del>



# Статистические методы обезличивания. Обобщение

## Округление

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	170	35	Рак
Артём	178	28	Грипп		*	180	30	Грипп
Иван	175	33	ОРВИ		*	175	35	ОРВИ



# Статистические методы обезличивания. Обобщение

## Нисходящее и восходящее кодирование

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	172	35	Рак
Артём	178	28	Грипп		*	175	30	Грипп
Иван	175	33	ОРВИ		*	175	35	ОРВИ



# Статистические методы обезличивания. Обобщение

## Комбинирование набора атрибутов в один атрибут

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	170-180	25-35	Рак
Артём	178	28	Грипп		*	170-180	25-35	Грипп
Иван	175	33	ОРВИ		*	170-180	25-35	ОРВИ



# Статистические методы обезличивания. Обобщение

## Локальное обобщение

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	170-180	25-35	Рак
Артём	220	28	Грипп		*	>200	25-35	Грипп
Иван	175	105	ОРВИ		*	170-180	>80	ОРВИ





# Статистические методы обезличивания. Искажение

## Метод добавления шума

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		*	168	34	Рак
Артём	178	28	Грипп		*	182	31	Грипп
Иван	175	33	ОРВИ		*	175	30	ОРВИ



# Статистические методы обезличивания. Искажение

## Метод замены/перестановки

Имя	Рост	Возраст	Болезнь		Имя	Рост	Возраст	Болезнь
Алексей	172	34	Рак		Петр	172	34	Рак
Артём	178	28	Грипп		Алексей	178	28	Грипп
Иван	175	33	ОРВИ		Артём	175	33	ОРВИ



# Статистические методы обезличивания. Искажение

## Агрегация или микроагрегация

Имя	Рост	Дата рождения		Имя	Рост	Дата рождения
Алексей	172	12.05.1990		Петр	172	1990
Артём	178	19.02.2000		Алексей	178	2000
Иван	175	01.01.1999		Артём	175	1999





# Статистические методы обезличивания. Перестановка

## Метрика I-разнообразия

Имя	Рост	Возраст	Зарплата		Имя	Рост	Возраст	Зарплата
Алексей	163	26	3		*	160-169	20-49	3
Артём	172	42	11		*	160-169	20-49	5
Тимофей	165	31	5	➔	*	170-179	20-49	3
Иван	181	38	5		*	170-179	20-49	11
Андрей	176	22	3		*	180-189	20-49	5
Максим	184	49	11		*	180-189	20-49	11
K = 1	L = 1				K = 2	L = 2		



# Статистические методы обезличивания. Перестановка

## Метрика t-близости

Имя	Рост	Возраст	Болезнь	Зарплата		Имя	Рост	Возраст	Болезнь	Зарплата
Алексей	163	26	Рак Желудка	3		*	160-174	20-29	Рак Желудка	3
Артём	172	28	Гастрит	4		*	160-174	20-29	Гастрит	4
Тимофей	165	24	Отравление	5		*	160-174	20-29	Отравление	5
Иван	181	42	ОРВИ	6		*	175-189	40-49	ОРВИ	6
Андрей	185	45	Рак Желудка	11	→	*	175-189	40-49	Рак Желудка	11
Максим	184	49	Аллергия	8		*	175-189	40-49	Аллергия	8
Илья	173	33	Грипп	7		*	160-174	30-39	Грипп	7
Пётр	169	30	ОРВИ	9		*	160-174	30-39	ОРВИ	9
Афанасий	171	39	Аллергия	10		*	160-174	30-39	Аллергия	10
K = 1	L = 1	T (Болезнь) = 1	T (Зарплата) = 1			K = 3	L = 3	T (Болезнь) = 0,5	T (Зарплата) = 0,375	

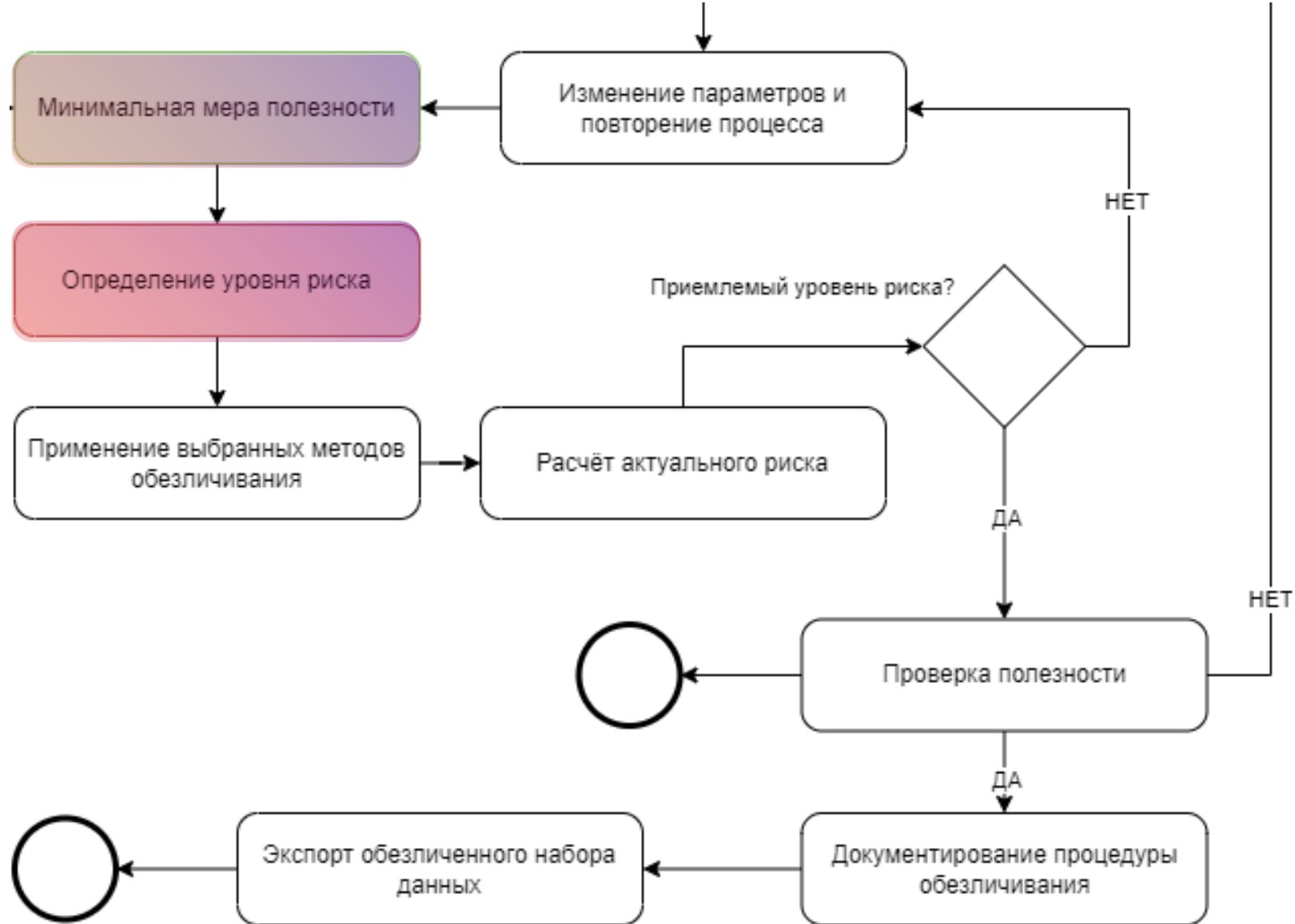


# Статистические методы обезличивания. Перестановка

## Метрика t-близости

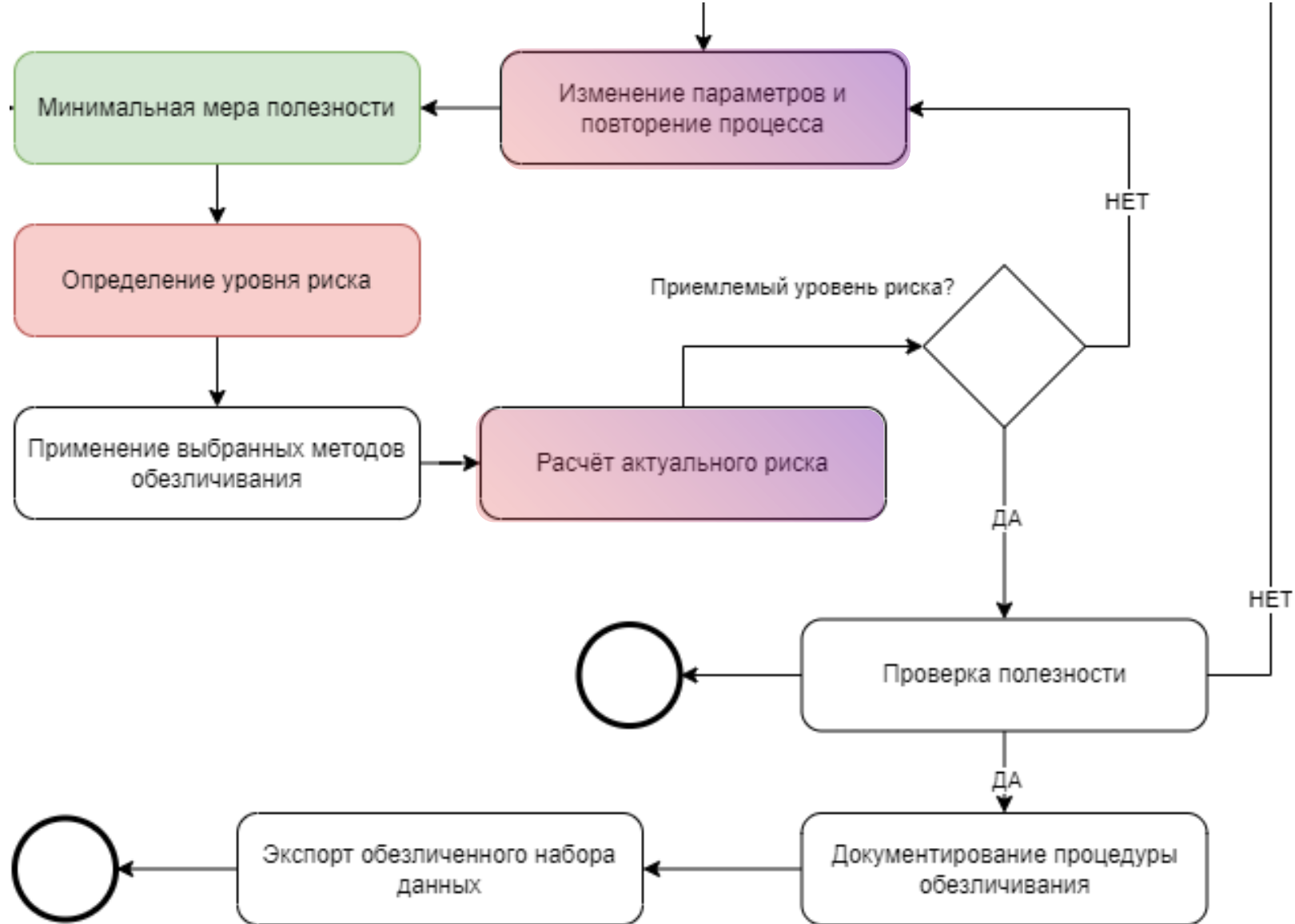
Имя	Рост	Возраст	Болезнь	Зарплата		Имя	Рост	Возраст	Болезнь	Зарплата
Алексей	163	26	Рак Желудка	3		*	160-169	20-39	Рак Желудка	3
Артём	172	28	Гастрит	4		*	160-169	20-39	ОРВИ	9
Тимофей	165	24	Отравление	5		*	160-169	20-39	Отравление	5
Иван	181	42	ОРВИ	6	→	*	180-189	40-60	ОРВИ	6
Андрей	185	45	Рак Желудка	11		*	180-189	40-60	Рак Желудка	11
Максим	184	49	Аллергия	8		*	180-189	40-60	Аллергия	8
Илья	173	33	Грипп	7		*	170-179	20-39	Грипп	7
Пётр	169	30	ОРВИ	9		*	170-179	20-39	Гастрит	4
Афанасий	171	39	Ковид	10		*	170-179	20-39	Ковид	10
K = 1	L = 1	T (Болезнь) = 1	T (Зарплата) = 1			K = 3	L = 3	0,278	T (Зарплата) = 0,167	

# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ

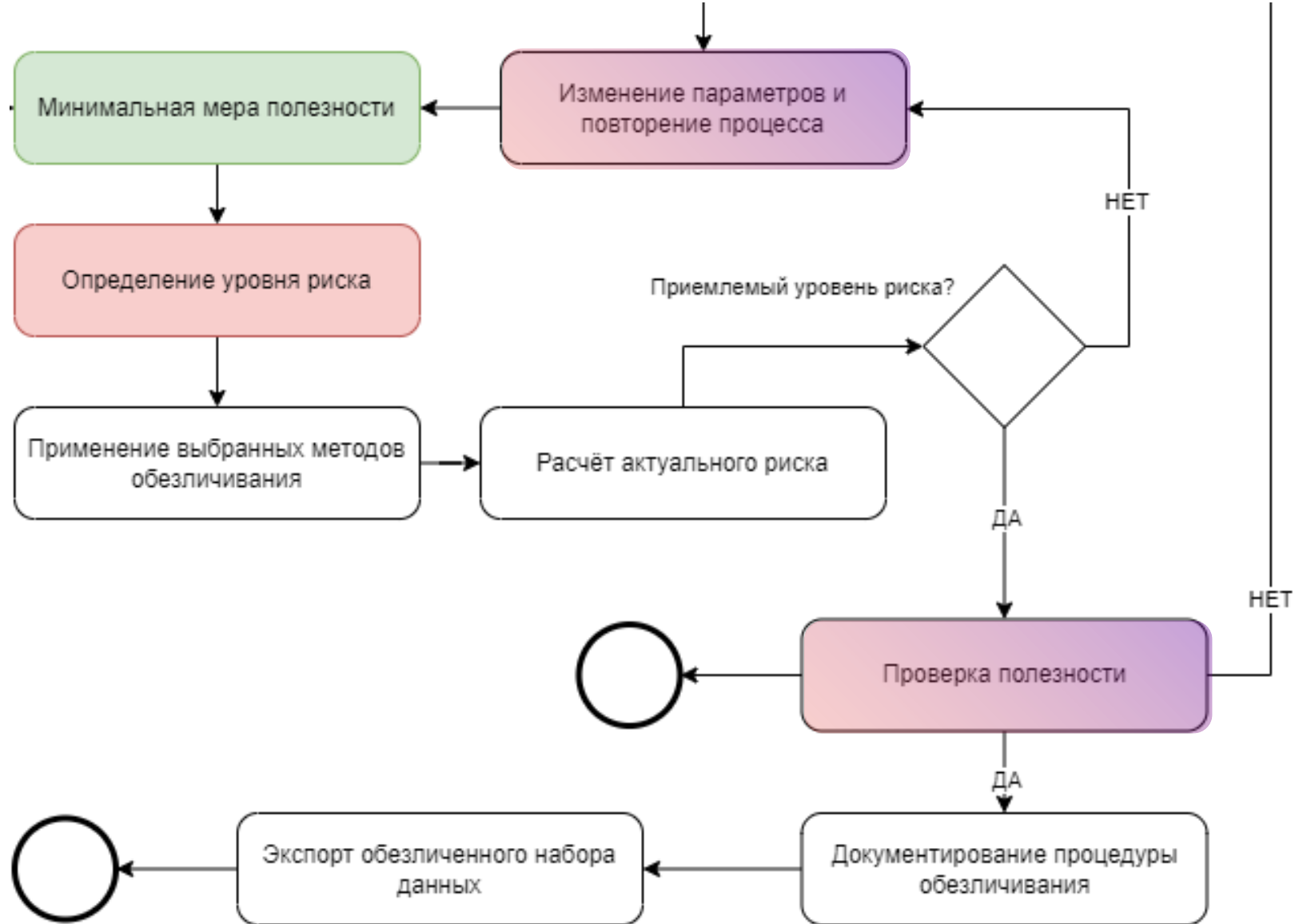




# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ



# МОДЕЛЬ ОЦЕНКИ РИСКОВ ДАННЫХ





# СУЩЕСТВУЮЩИЕ ИНСТРУМЕНТЫ

## Test Data Management and Dynamic Data Masking





# СУЩЕСТВУЮЩИЕ ИНСТРУМЕНТЫ

## Российское ПО для обезличивания



# СФЕРА.ОБЕЗЛИЧИВАНИЕ



АСОД v.15.08.2022 Главная Источники Приемники

Источники данных

Поиск

- Demo\_1
- Demo\_2
- Demo\_new

Главная > Источники данных > Demo\_new

## Demo\_new

1. Экспорт метаданных 2. Экспорт данных

Экспорт метаданных

Выгрузить

### Источник (изменение)

\* Наименование: Demo\_new

\* СУБД: POSTGRESQL

\* Строка подкл.: jdbc:postgresql://10.7.28.18:5432/source\_demo

\* Login: tdm\_reader

\* Пароль: .....

Тест подключения

\* Схемы: public

\* CI: 15555

Отмена Сохранить

7. Проверка

Подробнее Журнал



# СФЕРА.ОБЕЗЛИЧИВАНИЕ

АСОД v.15.08.2022 Главная Источники Приемники

Источники данных

Поиск

- Demo\_1
- Demo\_2
- Demo\_new

1. Экспорт метаданных 2. Экспорт данных 3. Профилирование 4. Анализ 5. Подготовка 6. Обезличивание 7. Проверка

Подробно Журнал

### Анализ

✓ Завершить анализ Скачать отчет

Схема Таблица Поле Тип Домен

Изменить отмеченные...

<input type="checkbox"/>	Схема	Таблица	Поле	Тип	Размер	Пустое	Вычисл.	Домен	Действие
<input type="checkbox"/>	public	client	client_id	int4	10	нет	нет	NO_PND	<a href="#">✎</a>
<input type="checkbox"/>	public	client	first_name	varchar	60	нет	нет	DMN_FIRST_NAME	<a href="#">✎</a>
<input type="checkbox"/>	public	client	middle_name	varchar	60	нет	нет	DMN_PATRONYMIC	<a href="#">✎</a>
<input type="checkbox"/>	public	client	last_name	varchar	60	нет	нет	DMN_SURNAME	<a href="#">✎</a>
<input type="checkbox"/>	public	client	birth_date	date	13	нет	нет	DMN_BIRTH_DATE	<a href="#">✎</a>
<input type="checkbox"/>	public	client	passport_nur	varchar	20	нет	нет	DMN_PASSPORT	<a href="#">✎</a>
<input type="checkbox"/>	public	client	passport_dat	date	13	нет	нет	DMN_DOC_DATE	<a href="#">✎</a>
<input type="checkbox"/>	public	client	address	varchar	255	нет	нет	DMN_ADDRESS	<a href="#">✎</a>
<input type="checkbox"/>	public	client	account	varchar	20	нет	нет	DMN_ACCOUNT	<a href="#">✎</a>
<input type="checkbox"/>	public	client	inn	varchar	4000	нет	нет	DMN_INN	<a href="#">✎</a>

# ГАРДА МАСКИРОВАНИЕ



Гарда Маскирование

Проекты

Текущий проект: DemoProject

Администратор

▶ Запустить маскирование

Новый проект

Исходная база данных | Конечная база данных | Настройки

Тип БД: Выберите элемент

Строка подключения: 255.255.255.255

Порт: Порт

Имя БД: Имя БД

Пользователь: Пользователь

Пароль: Пароль

Тестировать подключение

Сохранить

Отмена

Добавить проект

# ГАРДА МАСКИРОВАНИЕ



- Проекты
- Журнал заданий
- Настройки

МАСКИРОВАНИЕ ПРОЕКТА

- Сканирование
- Настройки маскирования**
- Шаблоны сканирования

Текущий проект: DemoProject

▶ Запустить маскирование

Схема БД Таблицы

База данных

Поиск по названию таблицы...

- ▶  DB1
- ▶  DB2
- ▶  DB3
- ▶  DB4
- ▶  DB5
- ▶  DB6

Все таблицы

Все данные Персональные Не персональные

Для маскирования подготовлено 486 колонок

Схема	Таблица	Колонка	Тип данных	ПДн	Вид маскирования	Пример
?	?	?	(Все)		(Все)	
DB1	Table1	Column1	int		Копирование	□
DB1	Table2	Column2	text		Копирование	□
DB1	Table3	Column3	date		Копирование	□
DB1	Table4	Column4	date		Копирование	□
DB1	Table5	Column5	int		Копирование	□
DB1	Table6	Column6	date		Копирование	□
DB1	Table7	Column7	text	AB	Перемешивание	□
DB1	Table8	Column8	text		Копирование	□
DB1	Table9	Column9	text		Копирование	□

Создать фильтр





# HF LABS МАСКИРОВЩИК

## Конфигурация маскировки

### Создать новую конфигурацию

Название конфигурации

Конфигурация-маскировки-демо1

#### Выберите файлы конфигурации

\*.yaml или \*.sql файл с конфигурацией маскирования

Выбрать файлы 

process\_main.yaml



process\_refresh.yaml



process\_delete.yaml



Создать конфигурацию 

### Готовые маскировки

#### клиенты\_Москва

process\_main.yaml

process\_refresh.yaml

 Connection\_test

 admin\_performer 12.12.2022 22:22:02

#### клиенты\_Питер

process\_main.yaml

process\_refresh.yaml

 Connection\_test

 admin\_performer 12.12.2022 22:22:02

#### партнеры\_Питер

process\_main.yaml

process\_refresh.yaml

process\_delete-all.yaml

[еще 9](#)

 Connection\_test

 admin\_performer 12.12.2022 22:22:02



# HF LABS МАСКИРОВЩИК



Мои маскировки

История маскировок

performer\_consent

< Назад к маскировкам

## Конфигурация-маскировки-демо1

session-config.yml

process\_refresh.yml

process\_delete.yml

+ Добавить файл

Настройка подключений

```
version: 0.1
description: Общий вид конфигурации «Маскировщика»

# Список БД, с которыми работаем в этом конфиге
databases:
  short_name_1:
    type: oracle
    connect_string: jdbc:oracle:thin:@localhost:1521:test
    schema: test_schema
    login: test_login
    password: 3314c96d411422d6ef5208f5b26fa5e4
    password_encoded: true
    pool_size: 50

# Обработчики (конфигурация)
handlers_configuration:
  common_fio_masking:
    params:
      processor_name: fio
      algorithm: mvp_fio
```

Файл process\_main.yml  
не содержит ошибок



# HF LABS МАСКИРОВЩИК



Мои маскировки

История маскировок

performer\_consent

< Назад к маскировкам

## Конфигурация-маскировки-демо1

session-config.yml

process\_refresh.yml

process\_delete.yml

+ Добавить файл

Настройка подключений

```
version: 0.1
description: Общий вид конфигурации «Маскировщика»

# Список БД, с которыми работаем в этом конфиге
databases:
  short_name_1:
    type: oracle
    connect_string: jdbc:oracle:thin:@localhost:1521:test
    schema: test_schema
    login: test_login
    password: 3314c96d411422d6ef5208f5b26fa5e4
    password_encoded: true
    pool_size: 50

# Обработчики (конфигурация)
handlers_configuration:
  common_fio_masking:
    params:
      processor_name: fio
      algorithm: mvp_fio
```

Файл process\_main.yml  
не содержит ошибок

# СРАВНЕНИЕ ИНСТРУМЕНТОВ



		Сфера.Обезличивание	Гарда.Маскирование	HFLabs Маскировщик	Сбер DeepDocs
<b>Поддерживаемые БД</b>	PostgreSQL	+	+	+	+
	Oracle	+	+	+	
	MS SQL	+	+		
	другие JDBC-источники	+	+	+	
	Hadoop	+			
	Elasticsearch				+
	doc, docx, txt, rtf, pdf, png, jpeg				+
	csv, json, xml	+			
<b>Настройка модели обучения</b>		+	+		
<b>Обезличенные данные реально выглядят</b>		+	+	+	+
<b>Сохраняются контрольные разряды и префиксы</b>		+	+	+	+
<b>Инкрементальная загрузка данных</b>		+	+	+	
<b>Сайт</b>		<a href="https://www.sferaplatform.ru/obezlichivanie-dannykh">wwhttps://www.sferaplatform.ru/obezlichivanie-dannykh</a> <a href="https://www.datprof.com">w.datprof.com</a>	<a href="https://gardatech.ru/products/gma/#functional">https://gardatech.ru/products/gma/#functional</a>	<a href="https://solutions.hflabs.ru/masking">https://solutions.hflabs.ru/masking</a>	<a href="https://www.sberbank.ru/promo/ecosystem/legal/anonymizer">https://www.sberbank.ru/promo/ecosystem/legal/anonymizer</a>



# NEOFLEX MASKIT

**Neoflex MaskIT**  
Depersonalization of data

## Sing In

Email or Login \*

Password \*



[Forgot password?](#)

Sign In

Create account



# NEOFLEX MASKIT

Neoflex MaskIT

Tasks History

Name Surname

Task info

Depersonalization

BD #1



Schema #1

Customer

Sales

Prices

Market index

Schema #2

History

Export

Hide the panel

Tasks / Task Name #1 / Depersonalization

## Task Name #1 / Depersonalization

Start

Add Connect

Customer

Columns	Source		Data domain	Degree of depersonalization	Depersonalization	
	Strings	Predict type			Strings	Predict type
Primary Key	16452	Integer	Primary key		12654	Integer
Name	Ruben	Varchar2	Name/Surn...		Jaydon	Varchar2
Surname	Arcand	Varchar2	Name/Surn...		*	Varchar2
Phone num	+7 999 666 77 88	Varchar2	Phone num...		+7 910 777 22 66	Varchar2
Address	Moscow Kremlin	Varchar2	Address		Moscow Arbat st.	Varchar2
Passport #	4277 333666	Numeric	Document...		4277 555123	Numeric
Sex	M	Varchar2	Sex		M	Varchar2
Date of birth	01.01.1956	Date	Date of birth		09.04.1974	Date
INN	77123456	Numeric	Document...		7753643	Numeric
Date of load	10.05.2013	Date	Date		12.08.2016	Date

**СПАСИБО  
ЗА ВНИМАНИЕ!**

Спикер:  
**Алексей Даньшин**