Большие данные — большая ответственность.

Опыт защиты от утечек в аналитических системах

Алексей Артемов
https://www.linkedin.com/in/aartemov/

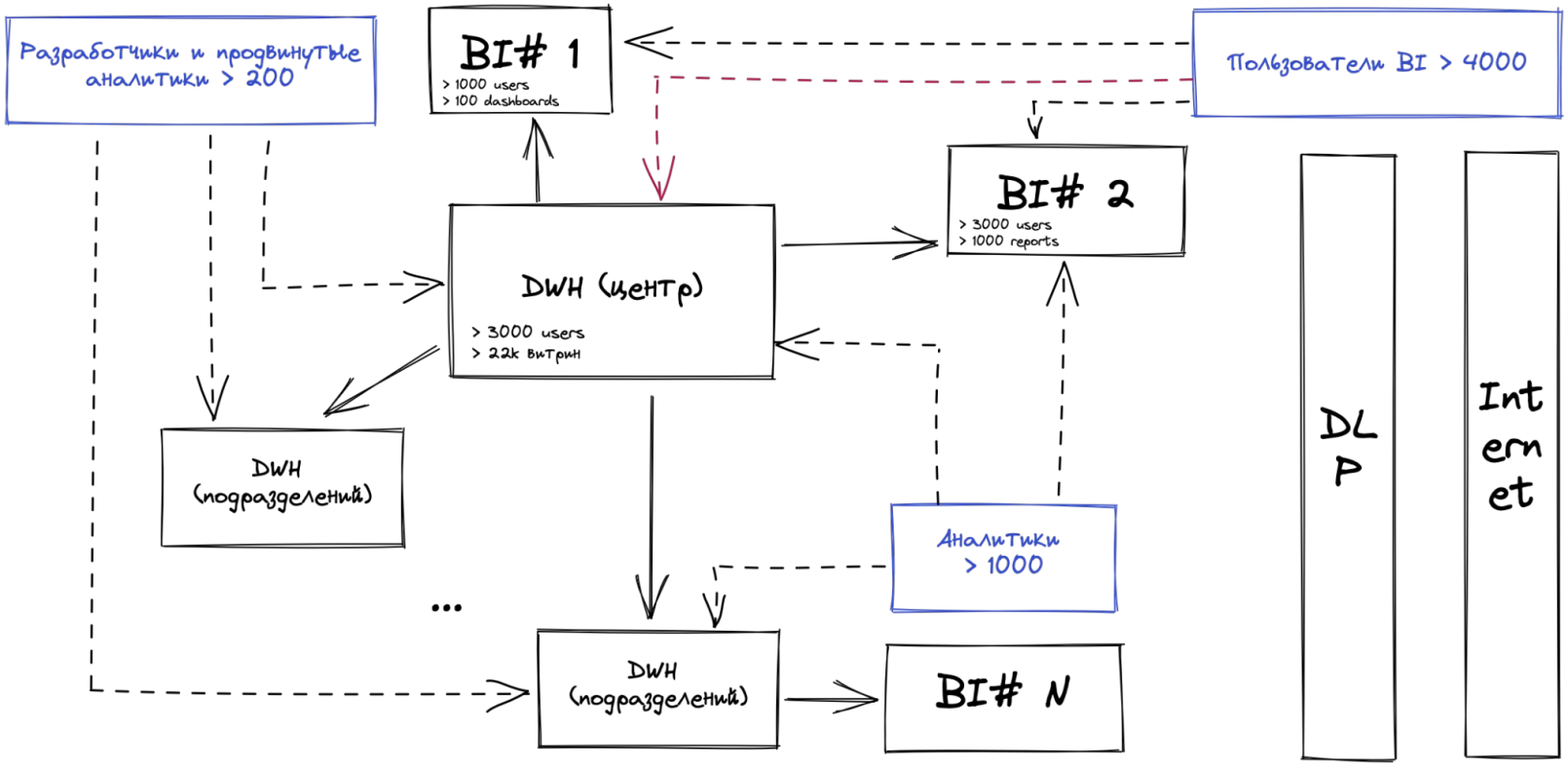
План

- Задача
- Исходная ситуация
- Вызовы
- Определение приоритетов
- Реализация
- Выученные уроки

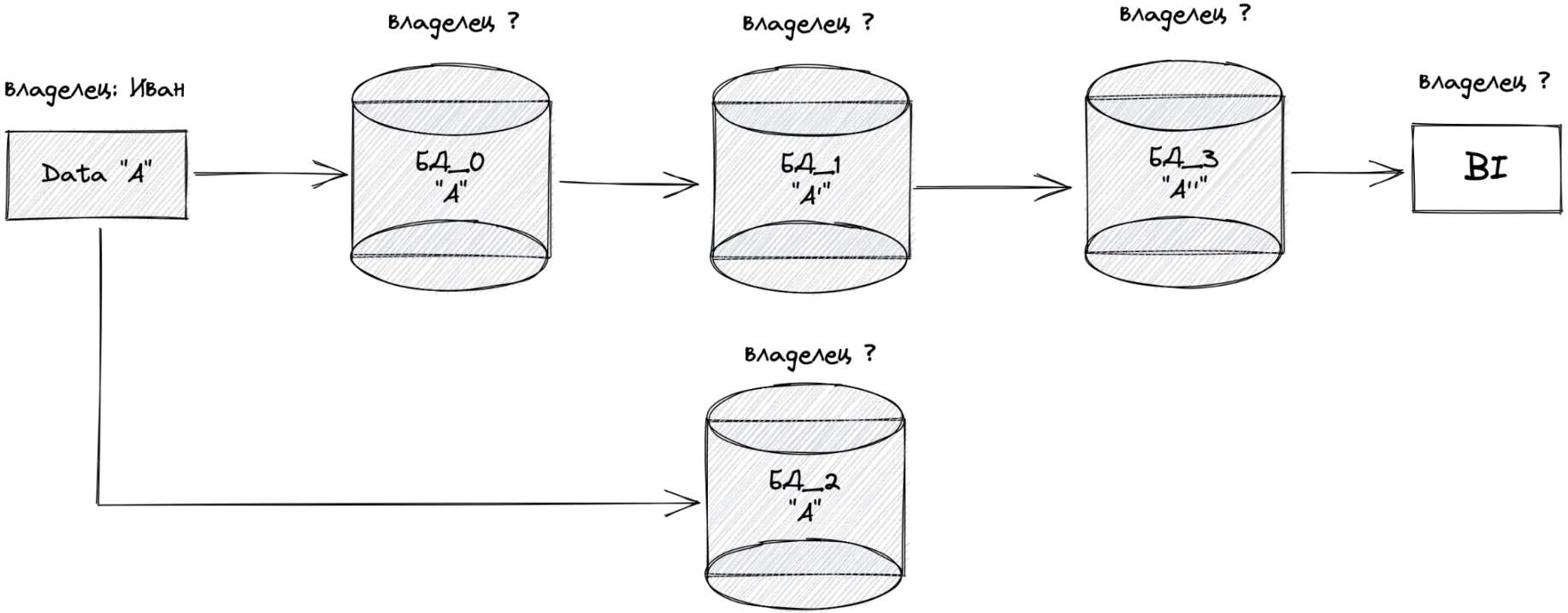
Задача

1. Необходимо было ограничить возможность неправомерной передачи данных 3м лицам
2. Не сломать работу компании
3. Сроки - ASAP

Исходная ситуация



Вызов 1 - “расползание” данных и ответственности



Вызов 2 - сложности с RLS & RBAC

1. RLS (row level security) - это здорово, но:
    a. Нужно знать к каким данным его необходимо применять
    b. Где-то вести привязку УЗ - Ключ справочника
    c. Большая работа на уже существующей инфраструктуре
2. RBAC (role based access control):
    a. Требует регулярной актуализации
    b. Потенциально большое кол-во ролей для управления

Вызов 3 - DLP

DLP система анализирует и перехватывает трафик компании для выявления конфиденциальной информации и, при необходимости, блокировки передачи данных.

Базовый принцип работы DLP систем – это фильтрация контента при отправке за периметр организации или в облако.

Исходная ситуация - итого

1. Доступ в Internet с рабочих станций
2. DWH уровня департаментов - Data Mesh?
3. “Песочницы” в DWH центр
4. Excel отчеты, которые разрабатывали сами департаменты
5. DLP - false-positive
6. Прямой доступ к DWH имели > 3000 пользователей
7. Нет реестра отчетов Excel и каталога данных

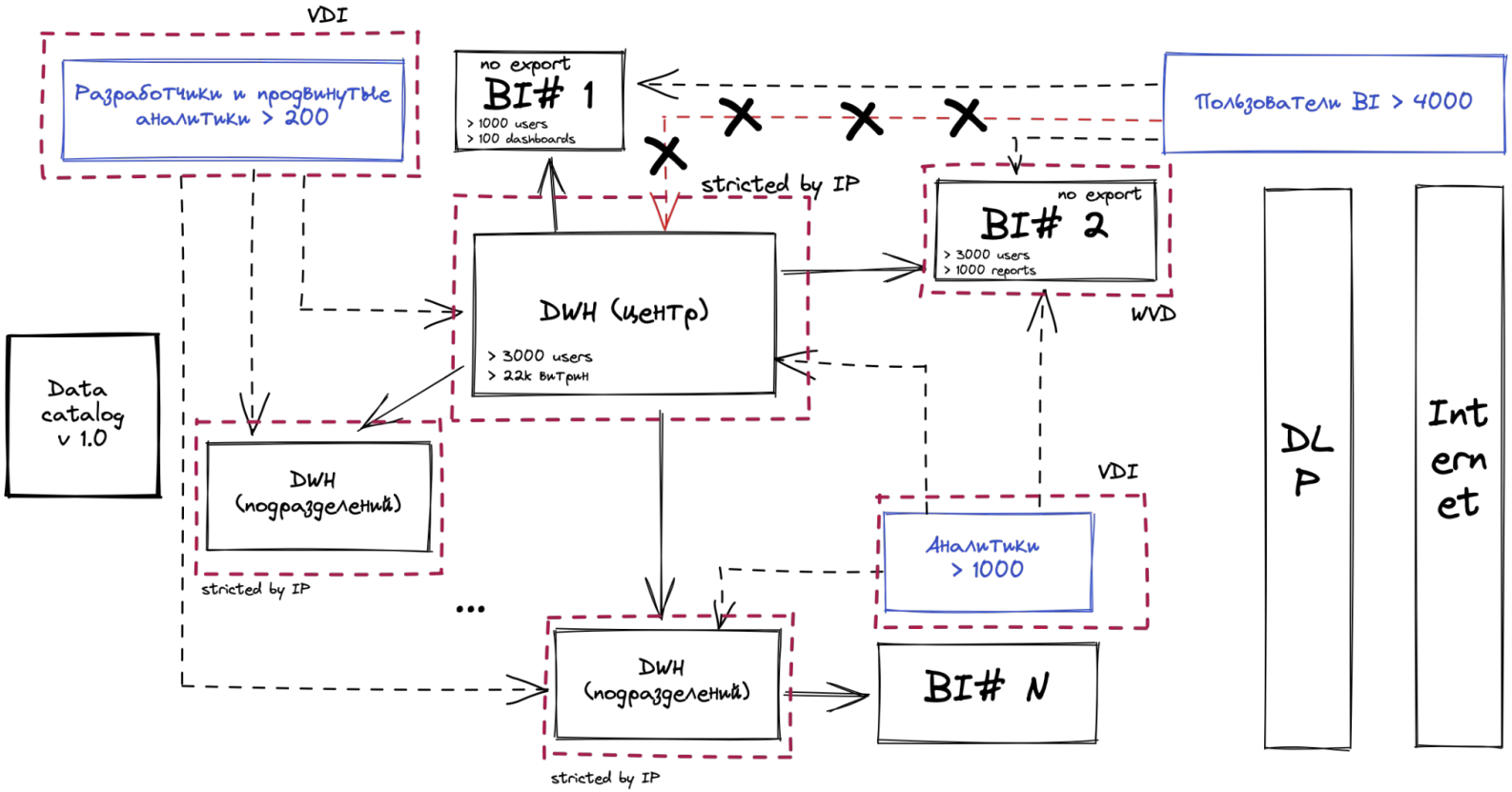
Определение приоритетов

1. Понять какие данные защищать → список систем подлежащих защите
2. Как защищать → оценка трудозатрат

Реализация

1. Убедили подразделения, что цель сохранить эффективность работы сотрудников
2. BI #1 → запрет экспорта данных
3. Excel → VD + RMS, с запретом экспорта данных
4. ~ 200 разработчиков → VDI
5. ~ 1000 аналитиков → VDI
6. DWH → ограничение доступа по IP
7. Аудит существующих доступов и регулярный мониторинг доступов
8. DLP → останавливала все файлы подписанные RMS

Реализация


VDI
Разработчики и продвинутые аналитики > 200
no export
BI# 1
> 1000 users
> 100 dashboards
Пользователи BI > 4000
stricted by IP
no export
BI# 2
> 3000 users
> 1000 reports
WVD
DWH (центр)
> 3000 users
> 22k витрин
Data catalog v 1.0
DWH (подразделений)
stricted by IP
VDI
Аналитики > 1000
...
DWH (подразделений)
BI# N
stricted by IP
DLP
Internet

Выученные уроки

1. Эффективное функционирование компании и ее сотрудников - это важнейший приоритет
2. При определении уровня конфиденциальности данных - не полагайтесь слепо на существующие политики
3. Если не знаете как оценить стоимость данных - сходите в профильные подразделения
4. Проводите регулярный аудит доступов
5. Имейте актуальный Data Catalog данных\отчетов\выгрузок\... Проводите регулярную сверку фактически существующих таблиц\витрин\отчетов\... с тем, что у вас в Data Catalog
6. Используйте инструменты представления данных, которые позволяют запретить экспорт и копирование данных
7. Проводите регулярное информирование какие данные являются критичными, а какие не очень по повышению грамотности сотрудников по обращению с данными, рассказывайте им какие данные являются критичными и как их нужно обрабатывать. Контролируйте соблюдение правил хранения и обработки критичных данных

Q&A