Под капотом Vulners

Из чего мы его собрали

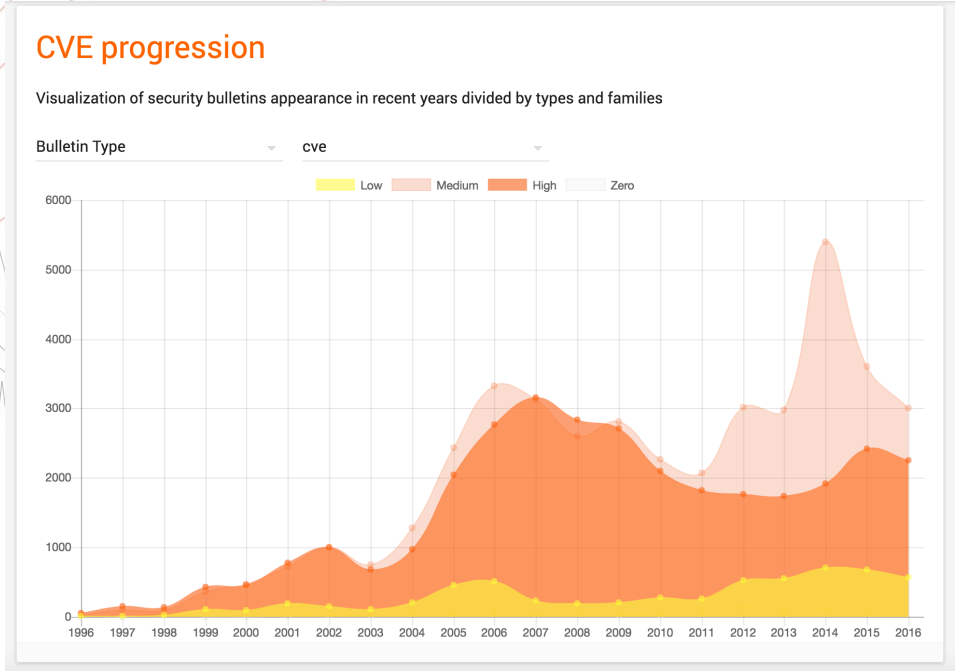
Кир Ермаков

Russian Open Source Summit 2017

#:ктоя

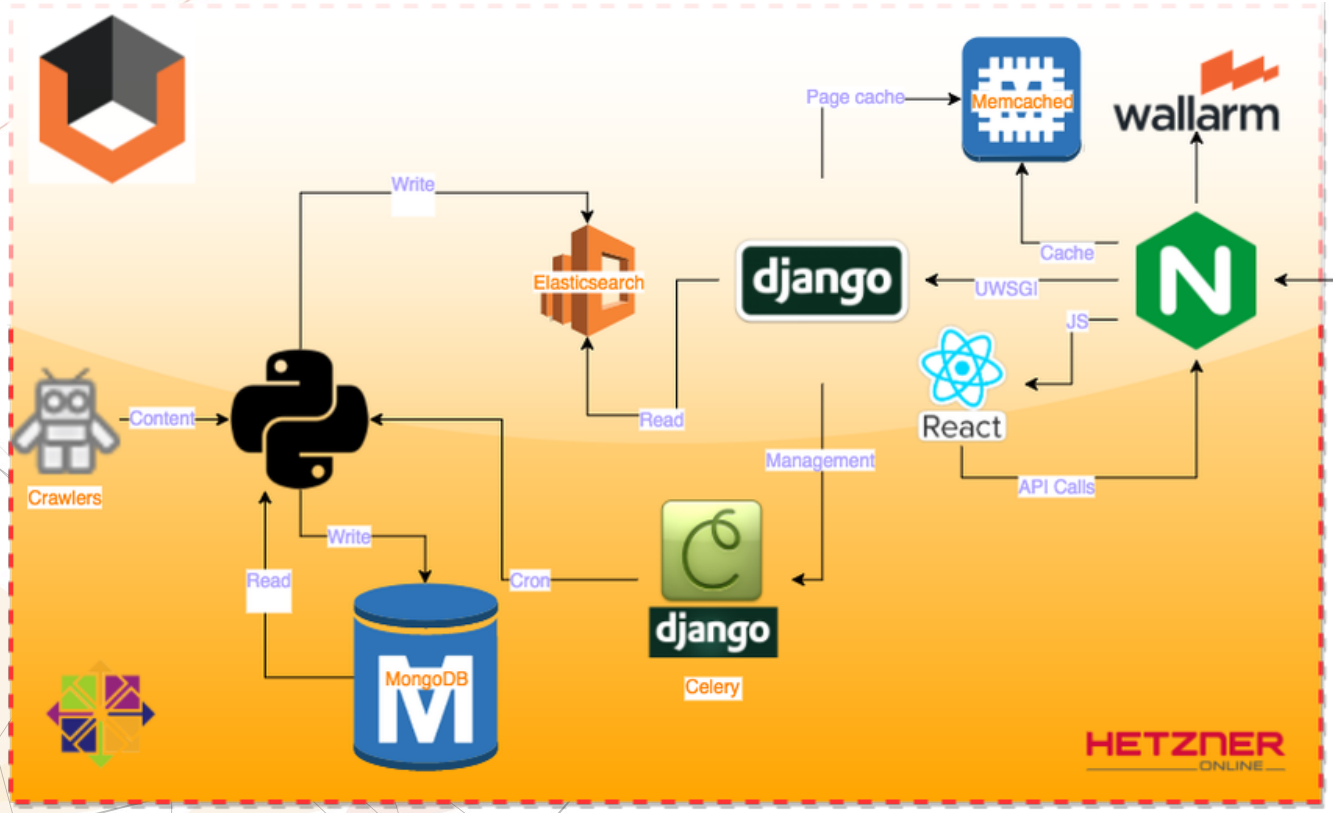
- Основатель vulners.com
- Технический директор группы QIWI
- Сайтики ломать умею
- В багбаунти неплохо так отметился (Yandex, Mail.ru, Apple…)
- Ярый активист «не_бумажной_безопасности»

База Данные Уязвимостей Vulners

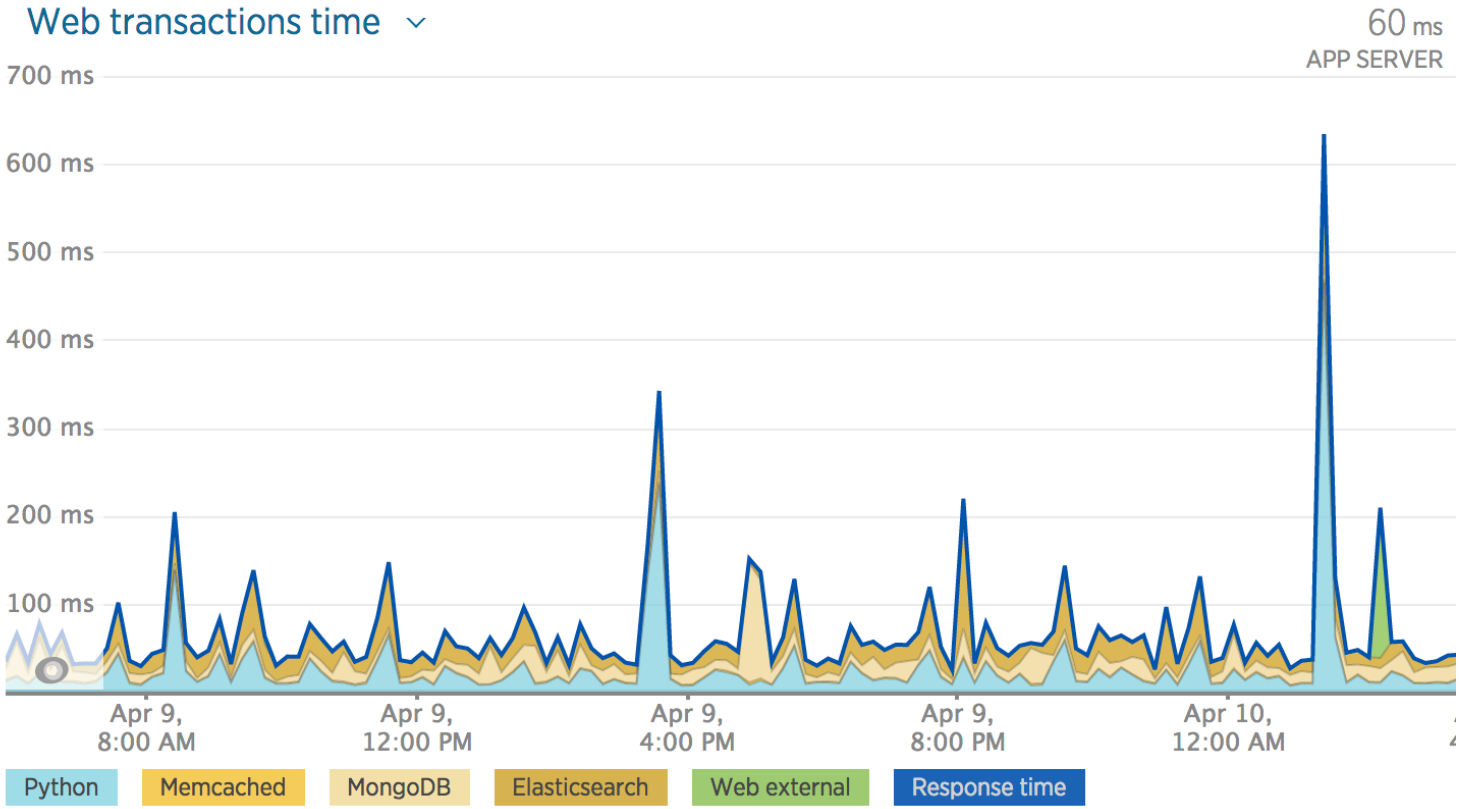


- Яндексо – подобный поисковик
- 600.000+ разнообразных элементов «по ИБэ»
- 71 источник данных
- Подписки и оповещения
- Умеет сканировать Линукс на уязвимости

Технологическая платформа



Шустро работает



Задняя часть

- Питон 3.6
  - Да, нам нравится асинхронность и юникод
- Джанго (не освобожденный, а фреймворк)
  - Со своими костылями и велосипедами
  - Собственная модель объектов, управление сеансами
- Рукастена(Wallarm), ДвижокИкс(Nginx) + UWSGI (непереводимо)
  - Интегрированная защита приложения и вполне обычный способ работы

requirements.txt

```
pyRSS2Gen
tweepy
pymongo
elasticsearch_dsl
elasticsearch
django-cors-headers
celery>4
django
telepot
passwordmeter
lepl
djangorestframework
pyparsing
bs4
recaptcha2
scrypt
argon2
six
pyga
version_utils
yandex.translate
pyopenssl
ndg-httpsclient
python3-memcached
whois
python-whois
django-pylibmc
django-mongodb-cash-backend
xhtml2pdf
cfscrape
markdown
weasyprint
pyrabbit
billiard
aiocfscrape
tqdm
yandex.translate
markdown2
celerybeat-mongo
```

Змеиная библиотека

- Ох и много их
- Обычный набор любителей Джанго
- Традиционное «а теперь разрули зависимости»

Планирование и краулинг

- Сельдерей 4.*
  - Подтягивает расписание и задачи из базы данных
- Кролик (очередь сообщений)
  - Просто очередь ☺
- Асинхронный вход/выход на Питоне
  - Что бы не расцеплять процесс краулинга на подпроцессы

Базы данных

- Гибкий поиск 2.*
  - Индекс, поиск и магия
- Монго Ди Би 3.*
  - Для внутренних объектов, долгоживущей сверхоперативной памяти и расписания планировщика
- Сверхоперативная-память-память (Memcache)
  - Веб-кеш и семафор с перекрестными процессами

Задержки

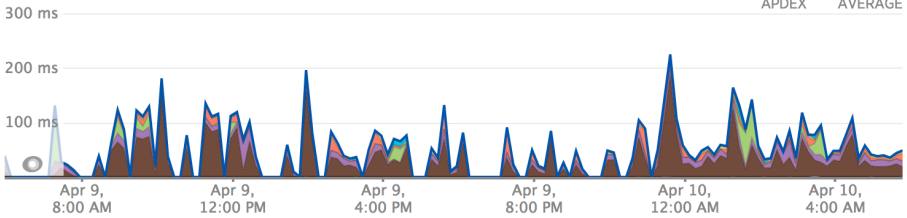
/search - lucene

App performance

App server breakdown

1.0 APDEX

72.7 ms AVERAGE



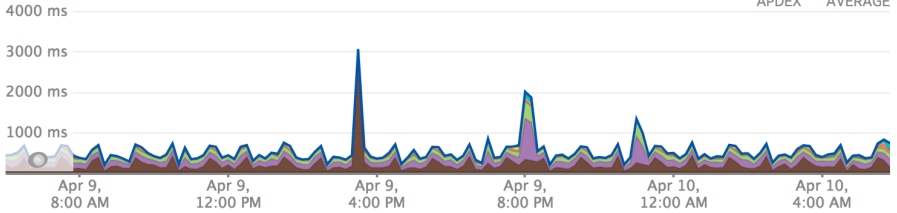
/audit - audit

App performance

App server breakdown

0.82 APDEX

618 ms AVERAGE



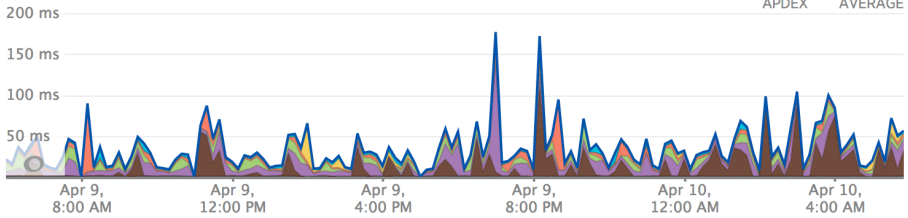
/search - id

App performance

App server breakdown

0.99 APDEX

39.6 ms AVERAGE



Кэш

- Обычный кэш на каждый день
  - Положим его в Мемкеш
- Плачущий вебкеш
  - Плачет из Мемкеша в Монго Ди Би
- Кеш для больших объектов
  - Файлы

Зачем столько баз?

- Исторически так сложилось
- Точный поиск гораздо менее больно делать в Монге
- Гибкий поиск что бы искать как Яндекс
- Сверхоперативная-память-память очень быстрая

Передний край

- Реактивный кофейный сценарий (ReactJS)
  - С кофе все можно делать быстрее
- Материальный интерфейс пользователя (MaterialUI)
  - Красивенько же!
- Веб-упаковщик(Webpack)
  - А теперь весь этот код надо собрать воедино
- ЙеЭс6 (ES6)
  - Нам нравится кофе с сахаром
  - Не поддерживает ослов. Ну и ладно.

Реактивный кофе скрипт == безопасность

Копируете контент с хакерских сайтов?

Нельзя просто так взять и вставить его в свой собственный сайт.

Чужой контент в нашем кофе не растворяется.

Вот так вот!

- Легкий
  - 4 центральных процессора + 32 гигабайта памяти
- Быстрый
  - <50мс на поисковые запросы
  - <500мс на сканирование пакетов
- Гибкий
  - Любой компонент умеет плодиться горизонтально

Спасибо!

- isox@vulners.com
- Продюсер попросил меня обязательно перевести слайды на русский язык
- Оригинальную презентацию я выложу на СлайдоПоделилке
- Стараемся делать мир лучше
- Хватит платить за то, что доступно бесплатно