



**IT NonStop**  
DataArt

Найти за полсекунды

---

Петр Петренко  
Senior software developer

skyeng



 valbok 7 февраля 2014 в 14:21

## Поиск изображений по фрагменту

Обработка изображений, Алгоритмы



# Схожие изображения

---



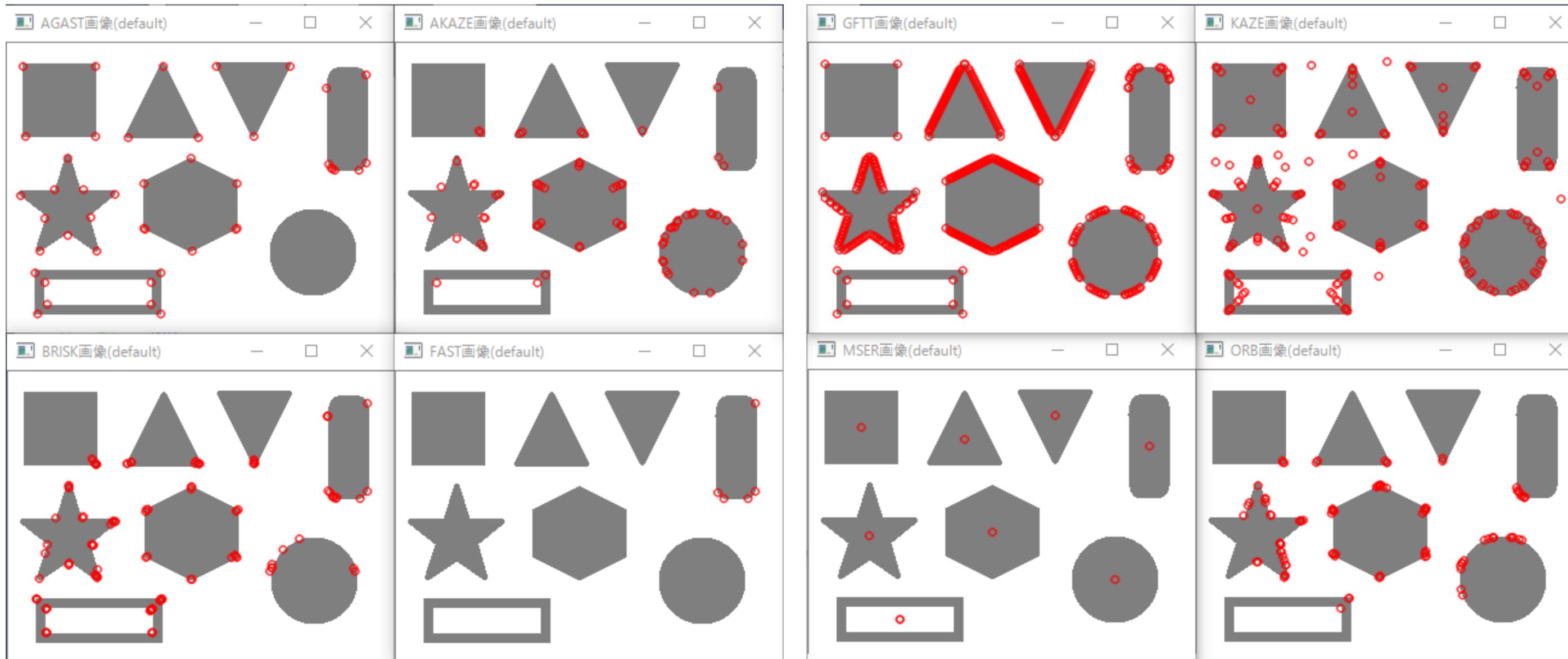
## Миниатюры



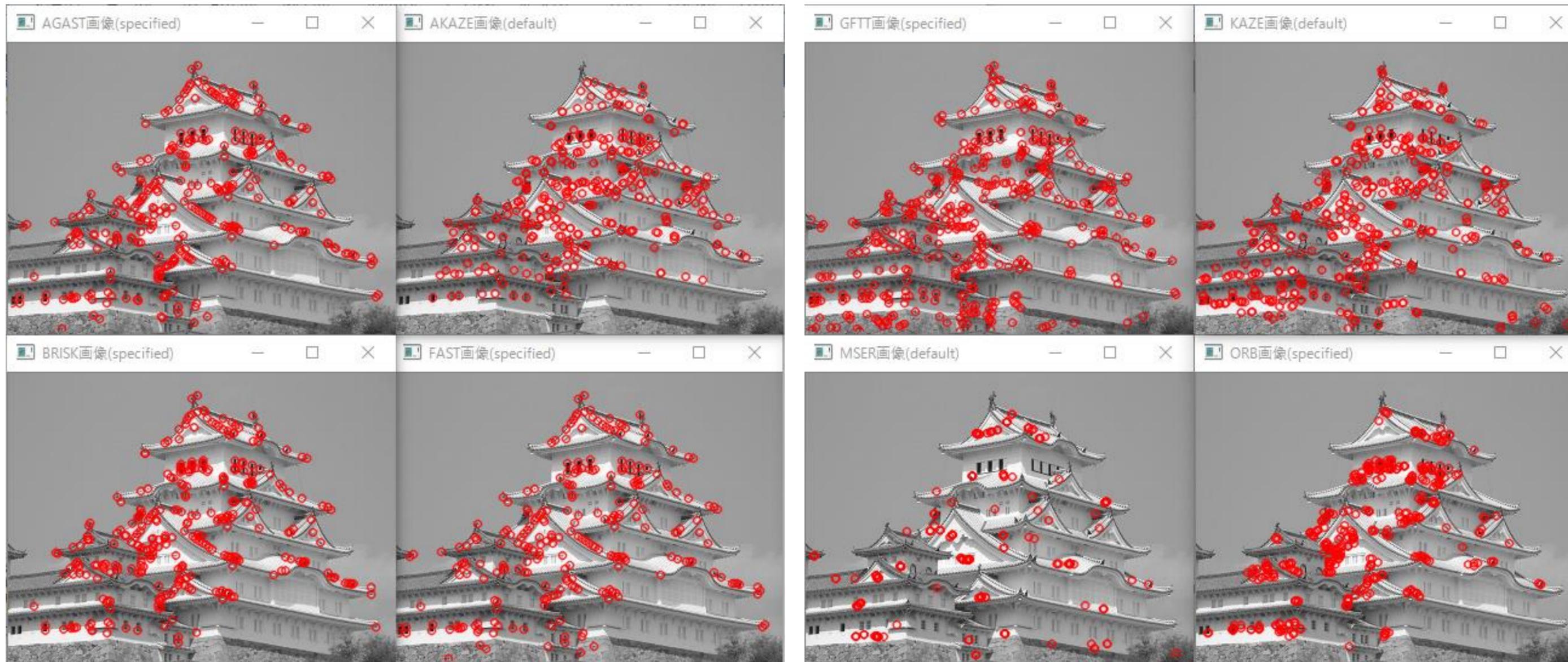
## Полудубликаты



# Ключевые точки



# Ключевые точки

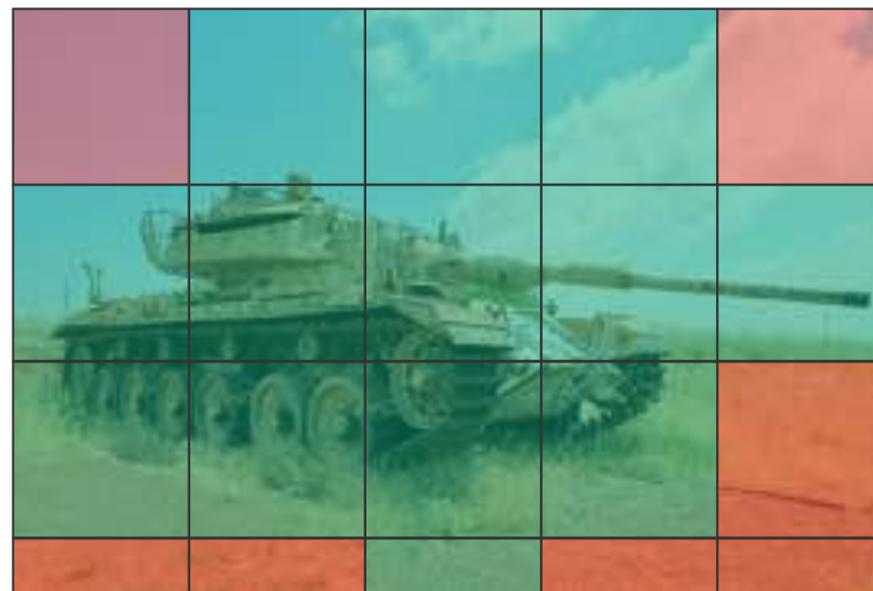


# Алгоритмы поиска ключевых точек



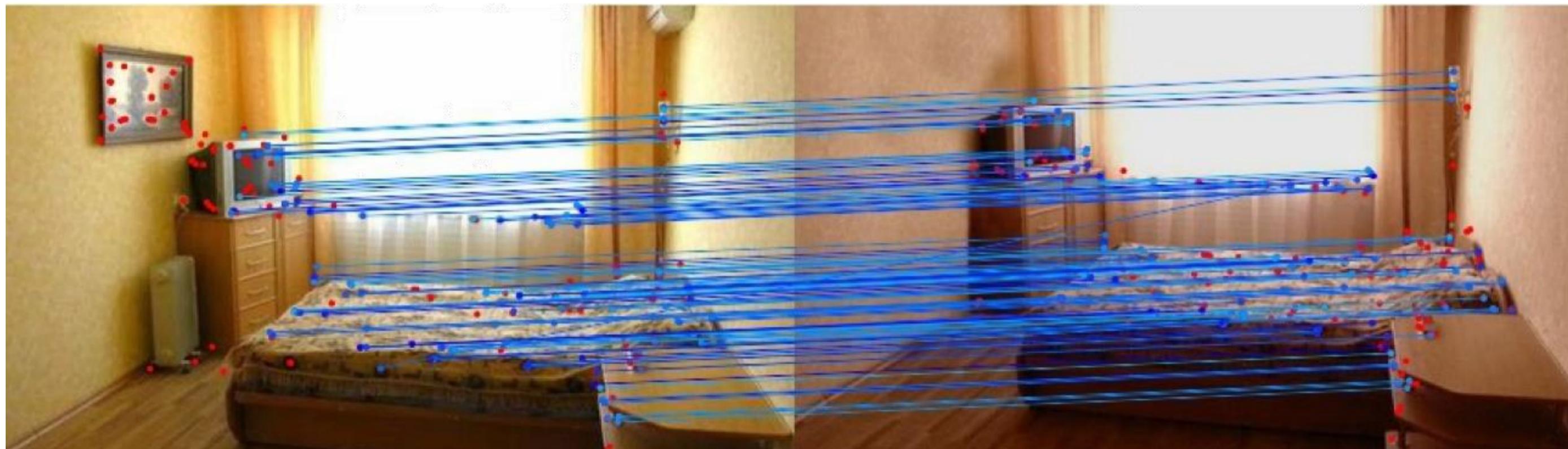
Алгоритм	Среднее покрытие	Максимальное покрытие	Минимальное покрытие	Среднее время выделения
FAST	81.60	237	4	<b><u>0.251 сек.</u></b>
BRISK	143.78	<b><u>367</u></b>	3	1.274 сек.
AKAZE	<b><u>155.57</u></b>	324	4	0.503 сек.

## Покрытие

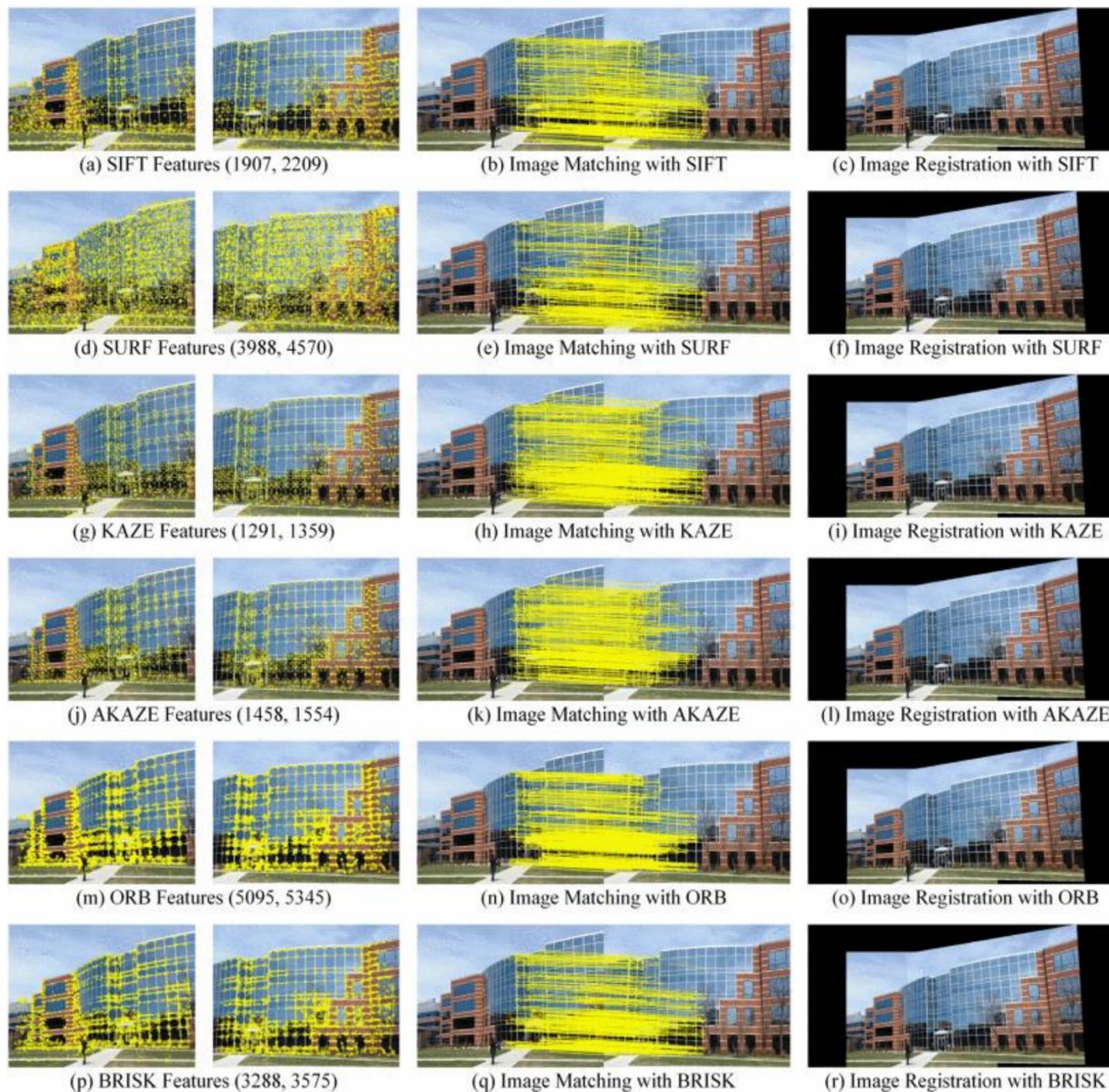


# Сопоставление точек

---

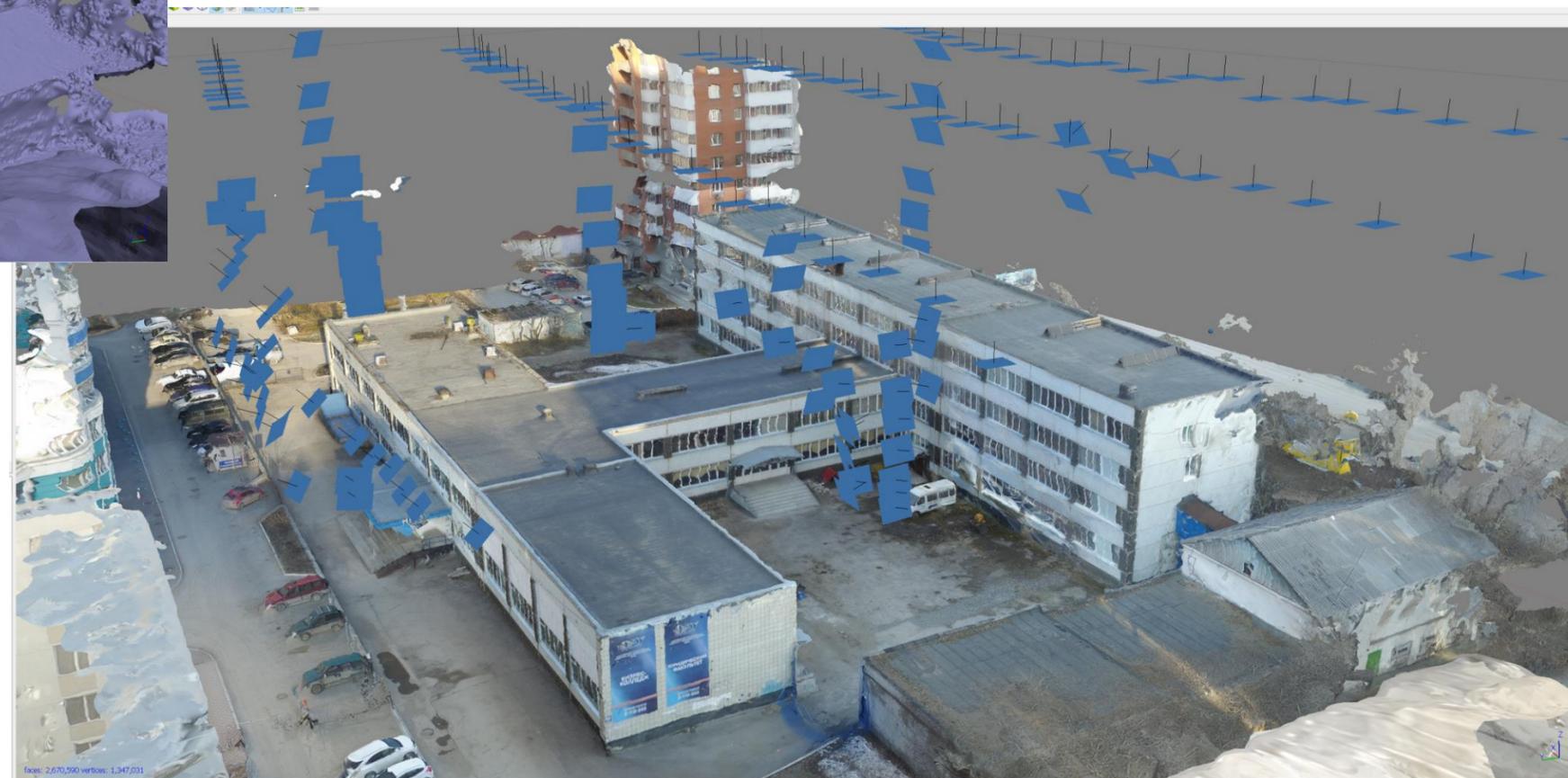
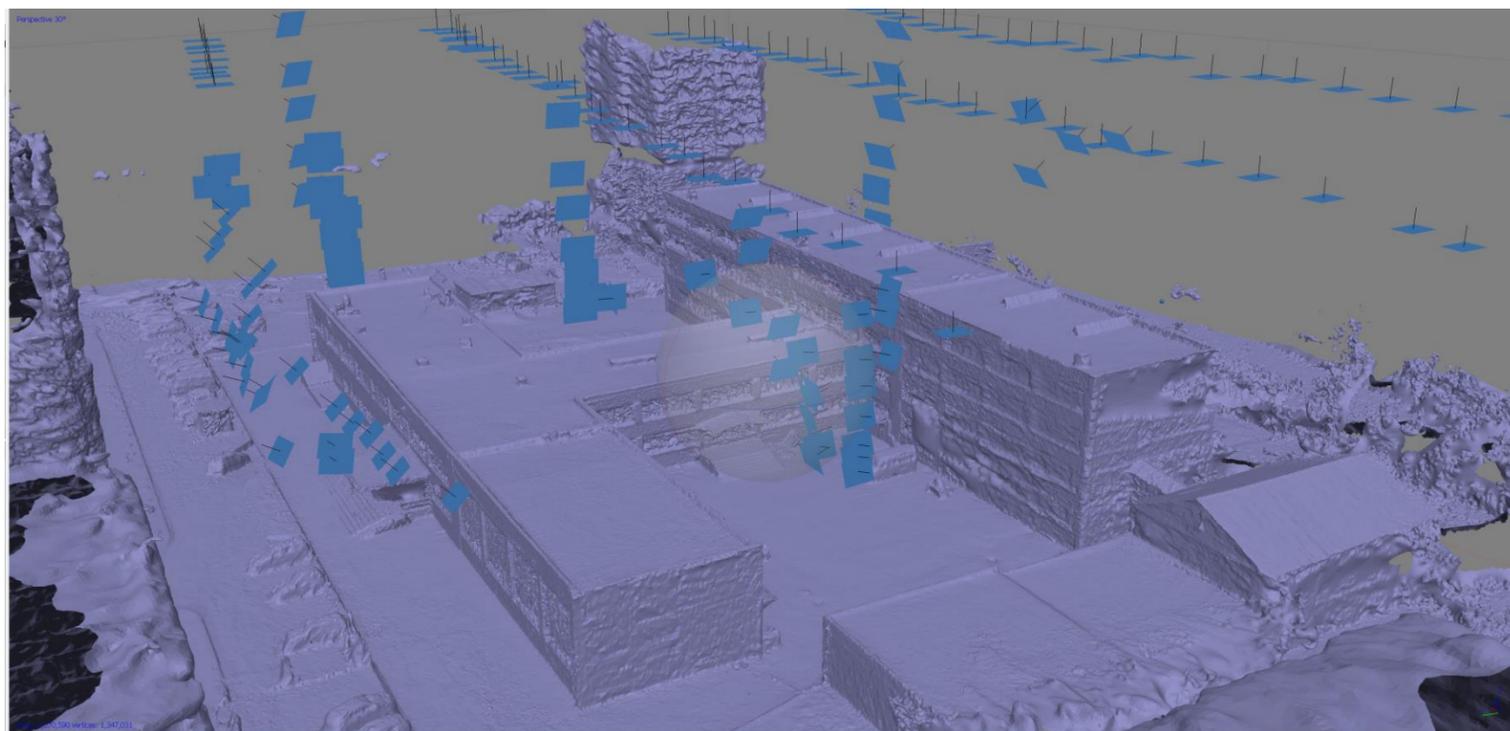


# Панорамы



# 3D модель по фото

---



# Дескрипторы



## BRISK: Descriptor

2D Gaussian around each feature

Robust to noise

Centers: **BLUE**    Gaussian: **RED**

32 slide: J. Heintz

(a) A patch with 3x3 gridding

(b) Binary test on a pair of grids

(c) A patch with multi-level gridding

G I    G II    G III

G IV    G V

BRISK

LDB

BRIEF

=> Feature vector (128)

SURF

# Сравнение дескрипторов



Рис 3.1. Сравнение уменьшенного в два раза изображения с оригиналом



Рис 3.3. Сравнение, обрезанного с разных сторон, изображения с оригиналом

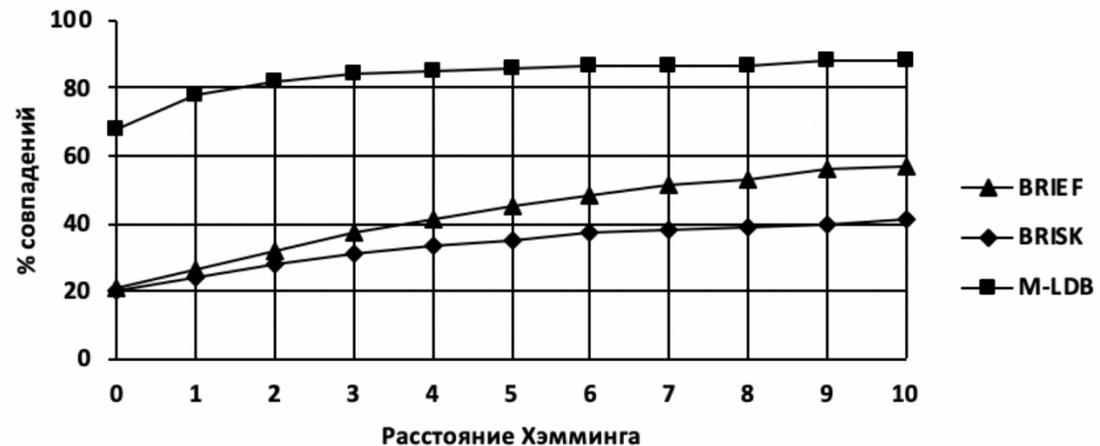


Рис 3.4. Сравнение, обрезанного снизу, изображения с оригиналом



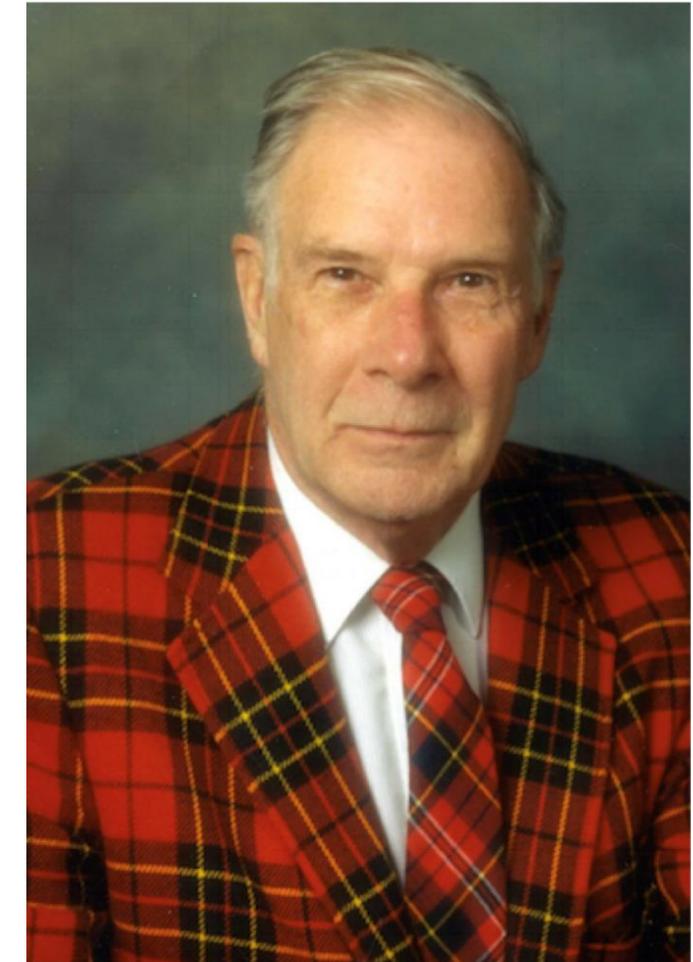
Рис 3.5. Сравнение разных изображений

# Расстояние Хэмминга

---



<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>1</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>



Число позиций, в которых соответствующие символы двух слов одинаковой длины различны

**Ричард Уэсли Хэмминг**

# Линейный поиск

---



Ключевых точек на одно изображение : ~**350**

Тестовая база данных : ~**350 миллионов точек**

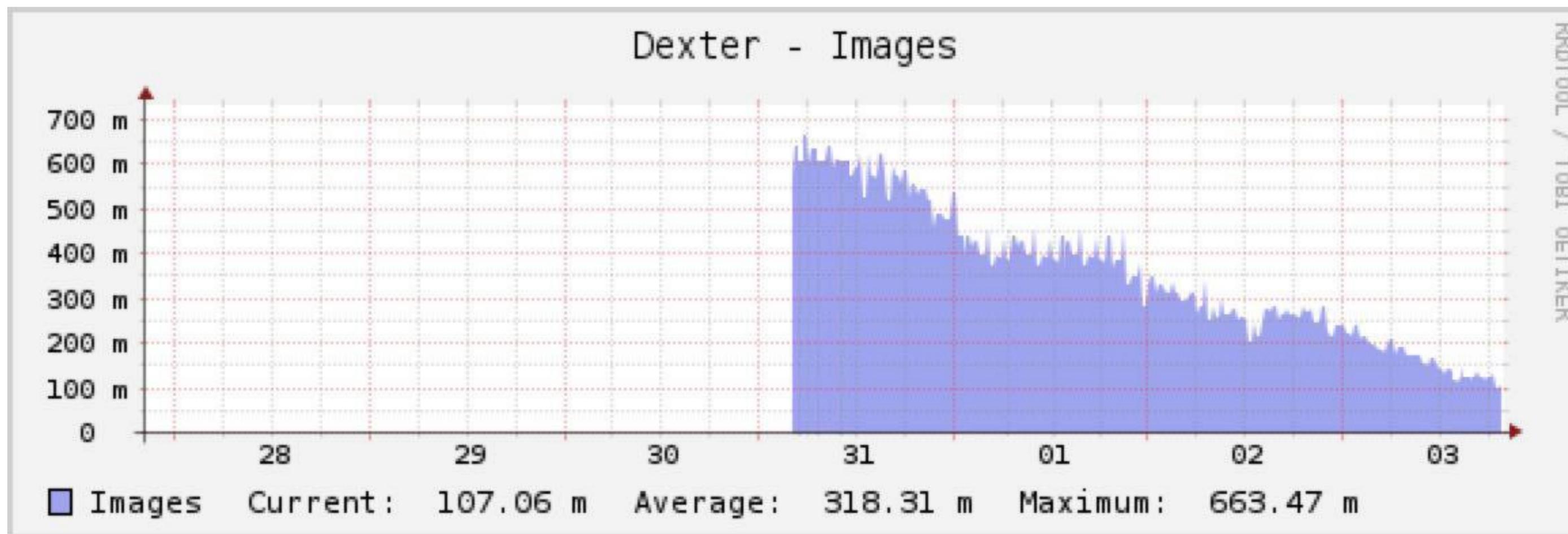
объем : ~ **10 гигабайт**

Линейный поиск одной ключевой точки в тестовой базе ~ **73.2 секунды** (без дисковых операций), всей картинке **более 7 часов**.

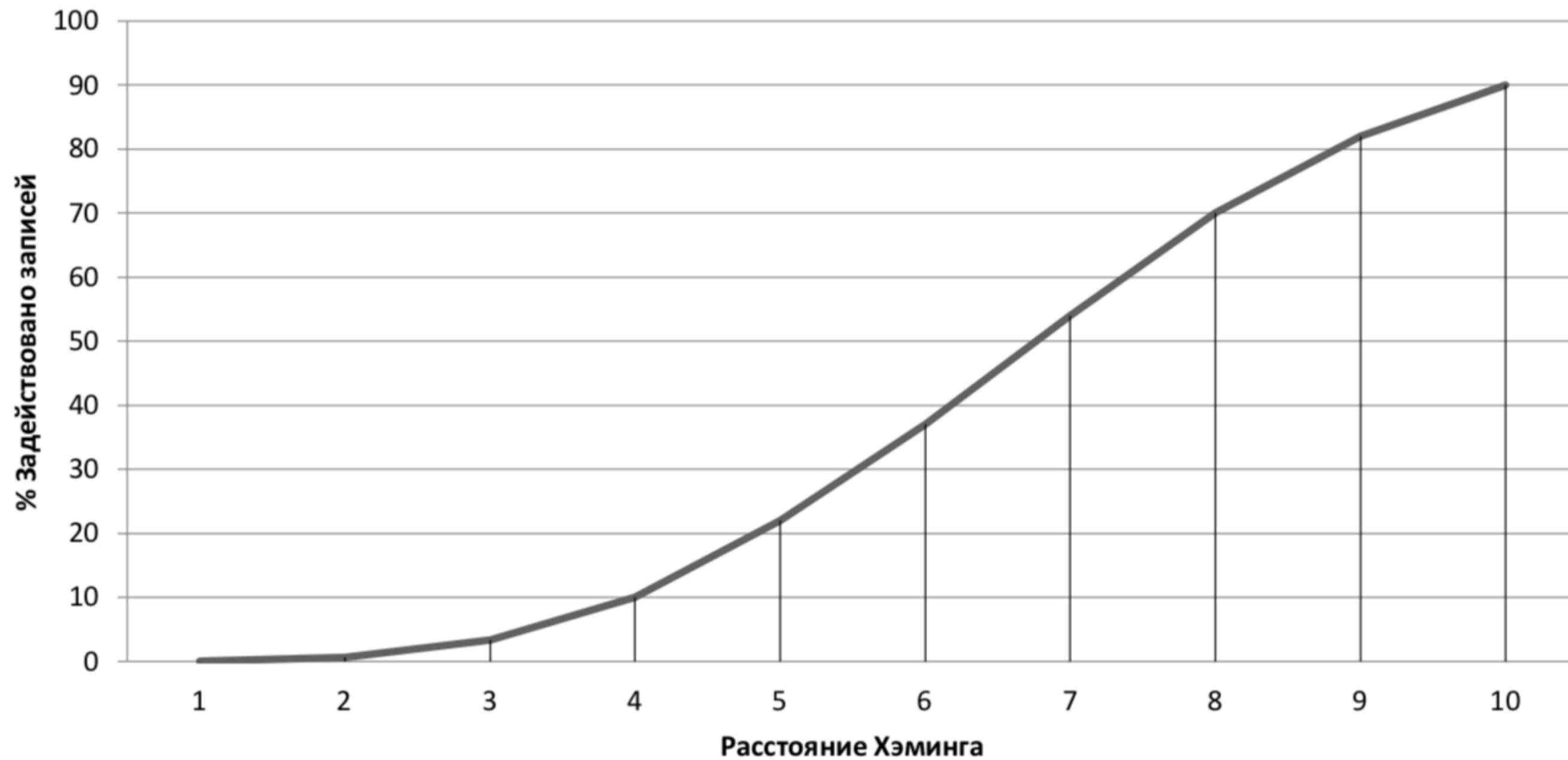
# HEngine



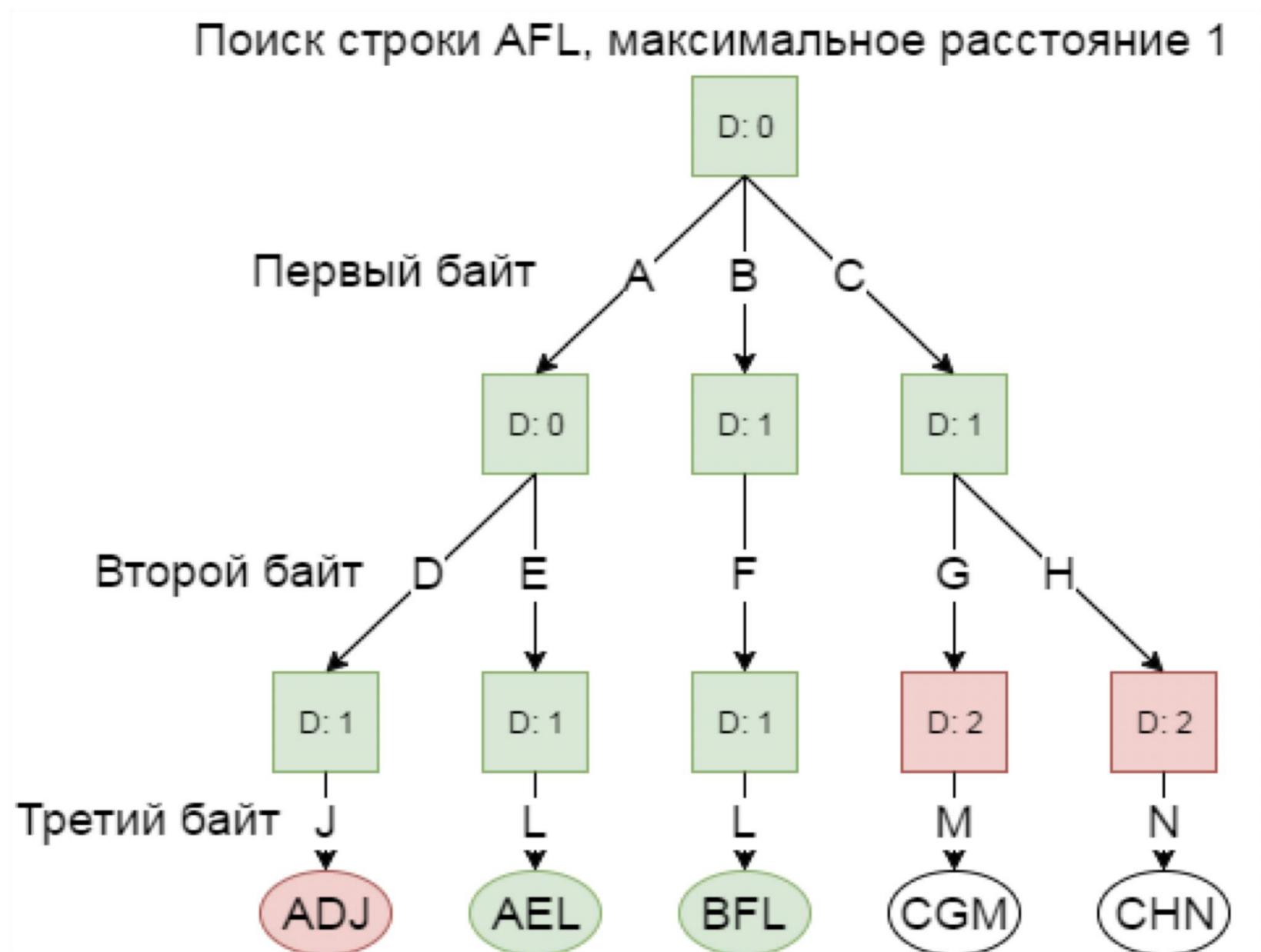
Объем базы данных: ~ 85 гигабайт



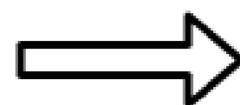
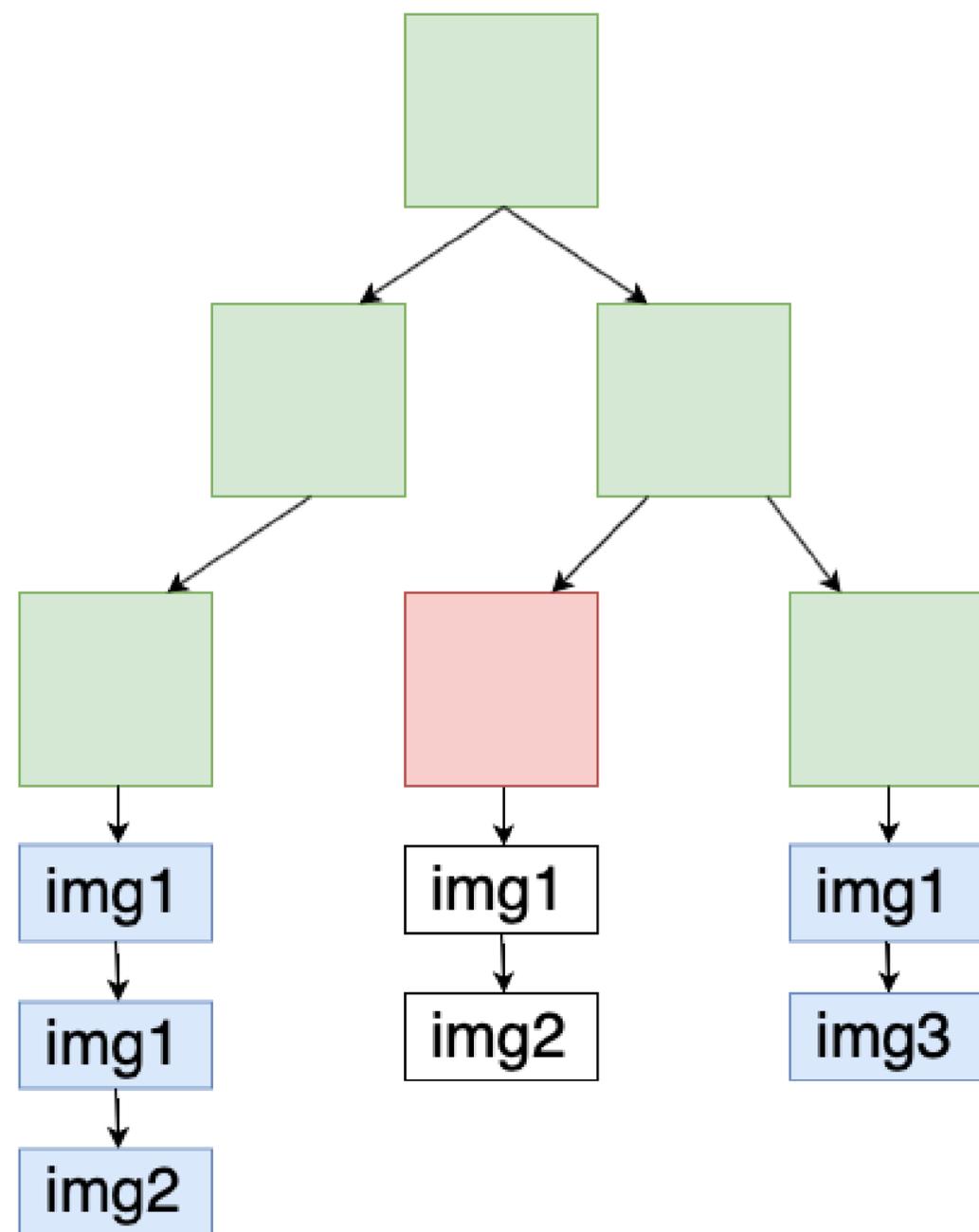
# VR-дерево



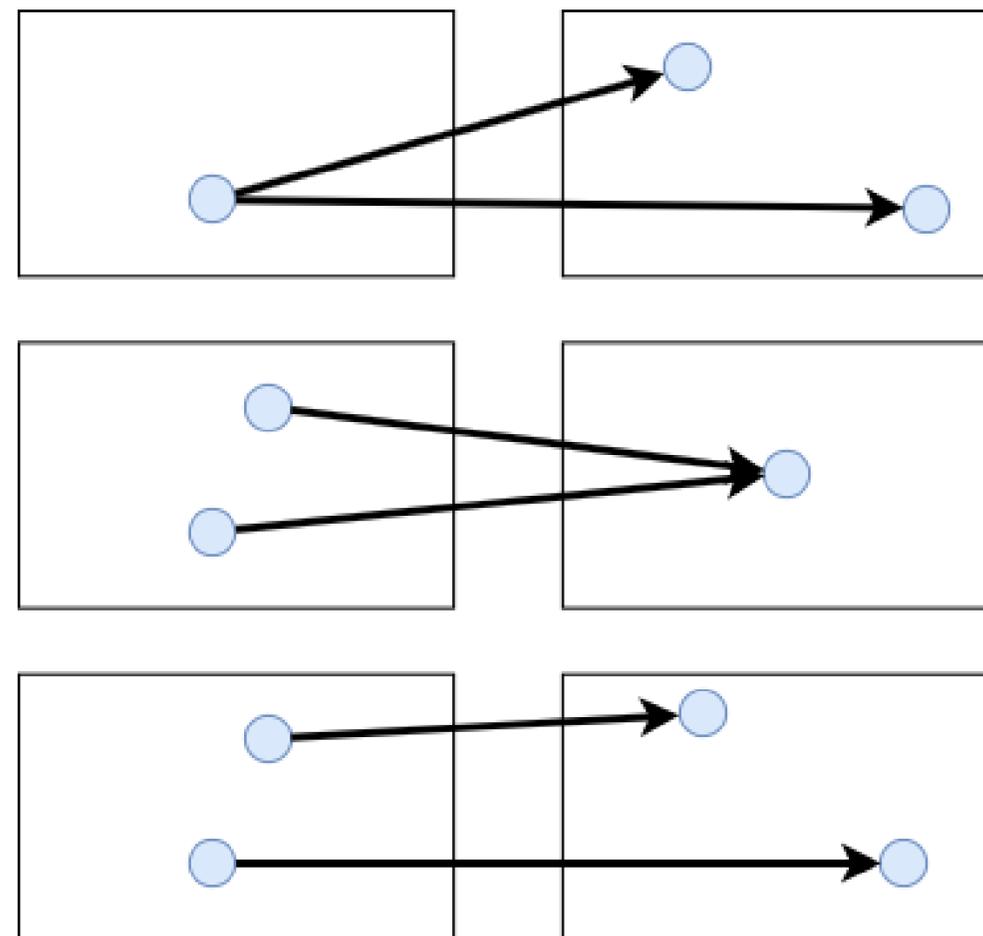
# Префиксное дерево



# Поиск



img1  
img2  
img3



# Оптимизация



Ключевых точек  $10 * 10^9$   
Узлов  $3 * 10^9$

		1 байт (среднее 3 записи) Количество потомков				
1 байт Размер префикса	N байт (среднее 5.8) Префикс	1 байт Тип записи	4 байт Размер записи	8 байт Ссылка	<b>32%</b> 19.8 байт	
1 байт Размер префикса	N байт (среднее 5.8 байт) Префикс	1 байт Тип записи	4 байт Размер массива	N * 4 байт (среднее 8 байт) Ссылка	<b>68%</b> 19.8 байт	

$19.8 * 3.3 + 1 = 66.34$  байта      **База 185 GB**

# Оптимизация



Ключевых точек $10 * 10^9$ Узлов $3 * 10^9$				
	1 байт (среднее 3 записи) Количество потомков			
1 байт Размер префикса	N байт (среднее 5.8) Префикс	1 байт Тип записи	4 байт Размер записи	8 байт Ссылка
1 байт Размер префикса	N байт (среднее 5.8 байт) Префикс	1 байт Тип записи	4 байт Размер массива	N * 4 байт (среднее 8 байт) Ссылка

**32% 19.8 байт**

**68% 19.8 байт**

$19.8 * 3.3 + 1 = 66.34$  байта    **База 185 GB**

Записей $10 * 10^9$ Узлов $3 * 10^9$					
		1 байт (среднее 3 записи) Количество потомков			
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8) Префикс	5 байт Ссылка		
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки		
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки	4 байта Id картинки	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки	4 байта Id картинки	4 байта Id картинки
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байт Размер массива	N * 4 байт (среднее 41.5 байт) Ссылка	

**32% 11.8 байт**

**55% 10.8 байт**

**8% 14.8 байт**

**2% 18.8 байт**

**3% 52.3 байт**

$12.8 * 3.3 + 1 = 43.24$  байта    **База 120 GB**

# Оптимизация



		1 байт (среднее 3 записи) Количество потомков	Записей $10 * 10^9$ Узлов $3 * 10^9$				
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8) Префикс	5 байт Ссылка			<b>32%</b> 11.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки			<b>55%</b> 10.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки	4 байта Id картинки			<b>8%</b> 14.8 байт
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байта Id картинки	4 байта Id картинки	4 байта Id картинки	<b>2%</b> 18.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	4 байт Размер массива	N * 4 байт (среднее 41.5 байт) Ссылка		<b>3%</b> 52.3 байт	

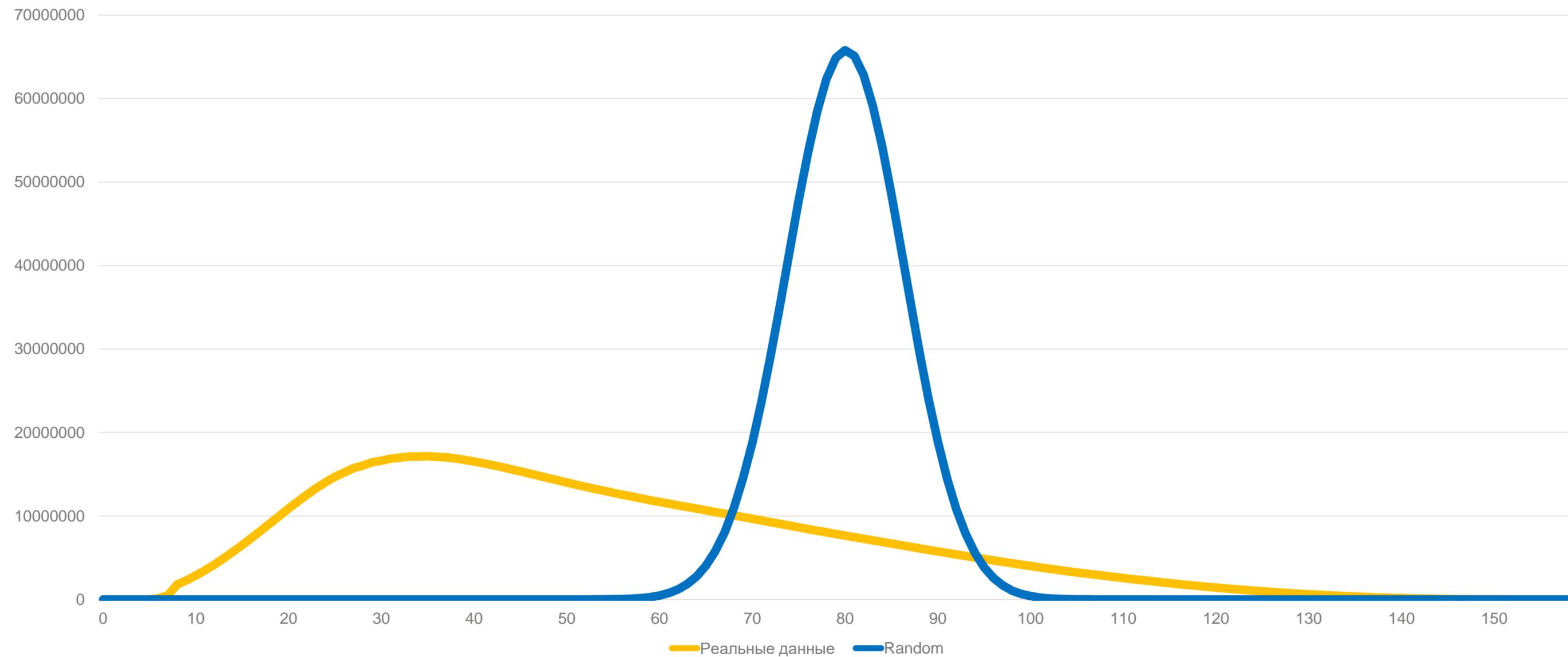
$12.8 * 3.3 + 1 = 43.24$  байта **База 120 GB**

		1 байт (среднее 3 записи) Количество потомков	Записей $10 * 10^9$				
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8) Префикс	5 байт Ссылка			<b>32%</b> 11.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	3 байта Id картинки			<b>55%</b> 9.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	3 байта Id картинки	3 байта Id картинки			<b>8%</b> 12.8 байт
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	3 байта Id картинки	3 байта Id картинки	3 байта Id картинки	<b>2%</b> 15.8 байт	
3 бита Тип записи	5 бит Размер префикса	N байт (среднее 5.8 байт) Префикс	2 байт Размер массива	N * 3 байт (среднее 31.1 байт) Ссылка		<b>3%</b> 39.9 байт	

$11.7 * 3.3 + 1 = 39.61$  байта **База 110 GB**

**-75 GB**

# Неравномерное распределение



# Префиксное дерево



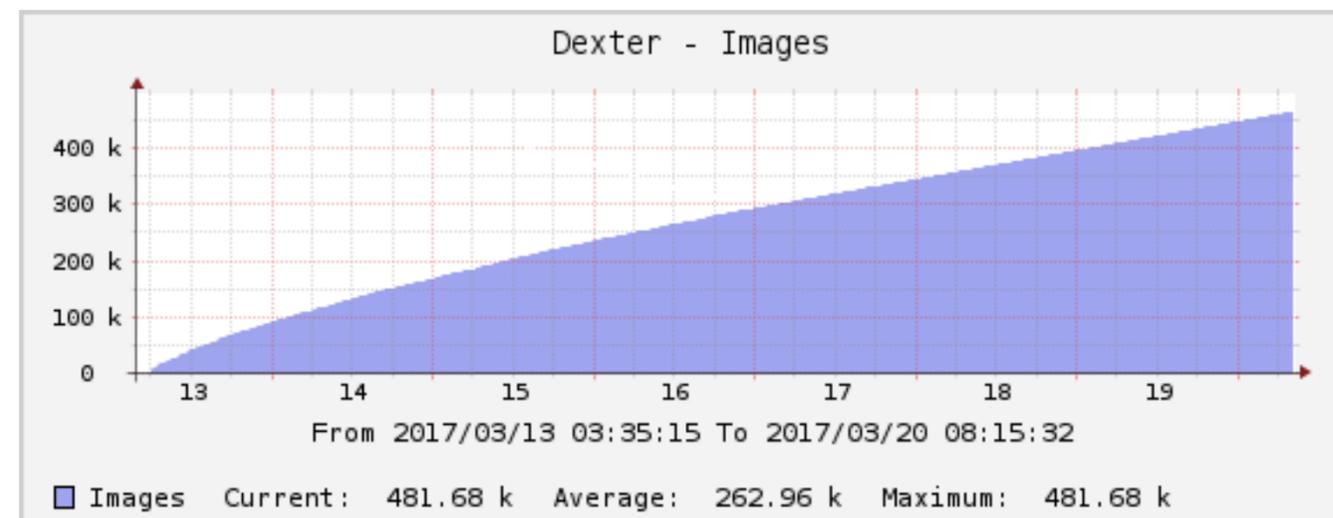
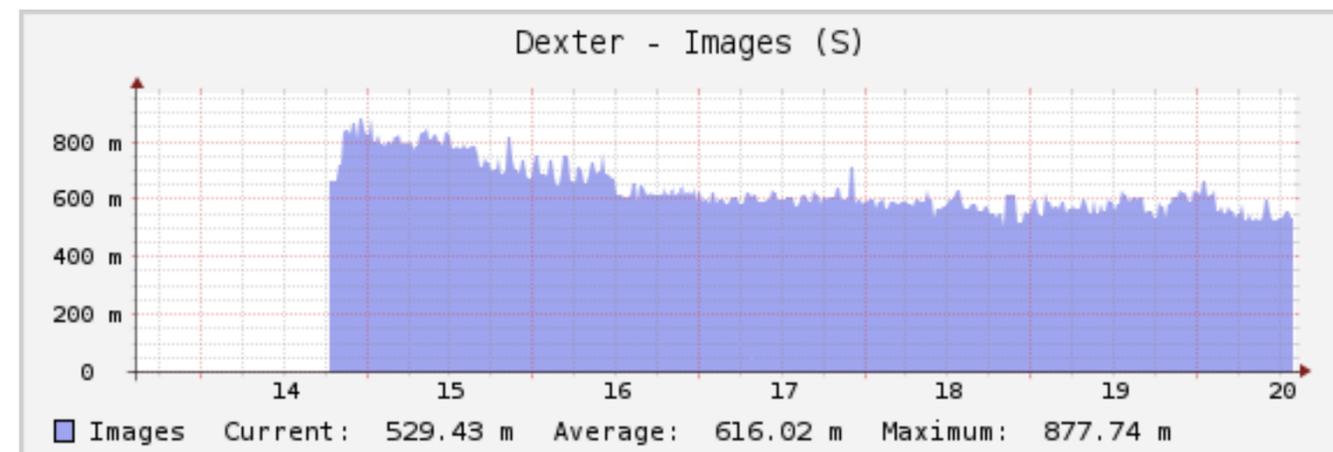
Объем файла с деревом: ~ 7 гигабайт

Поиск одной ключевой точки в тестовой базе примерно **3.5 миллисекунды**, всей картинки примерно **1.3 секунда**.

$1 * 10^6 - 7$  GB/M

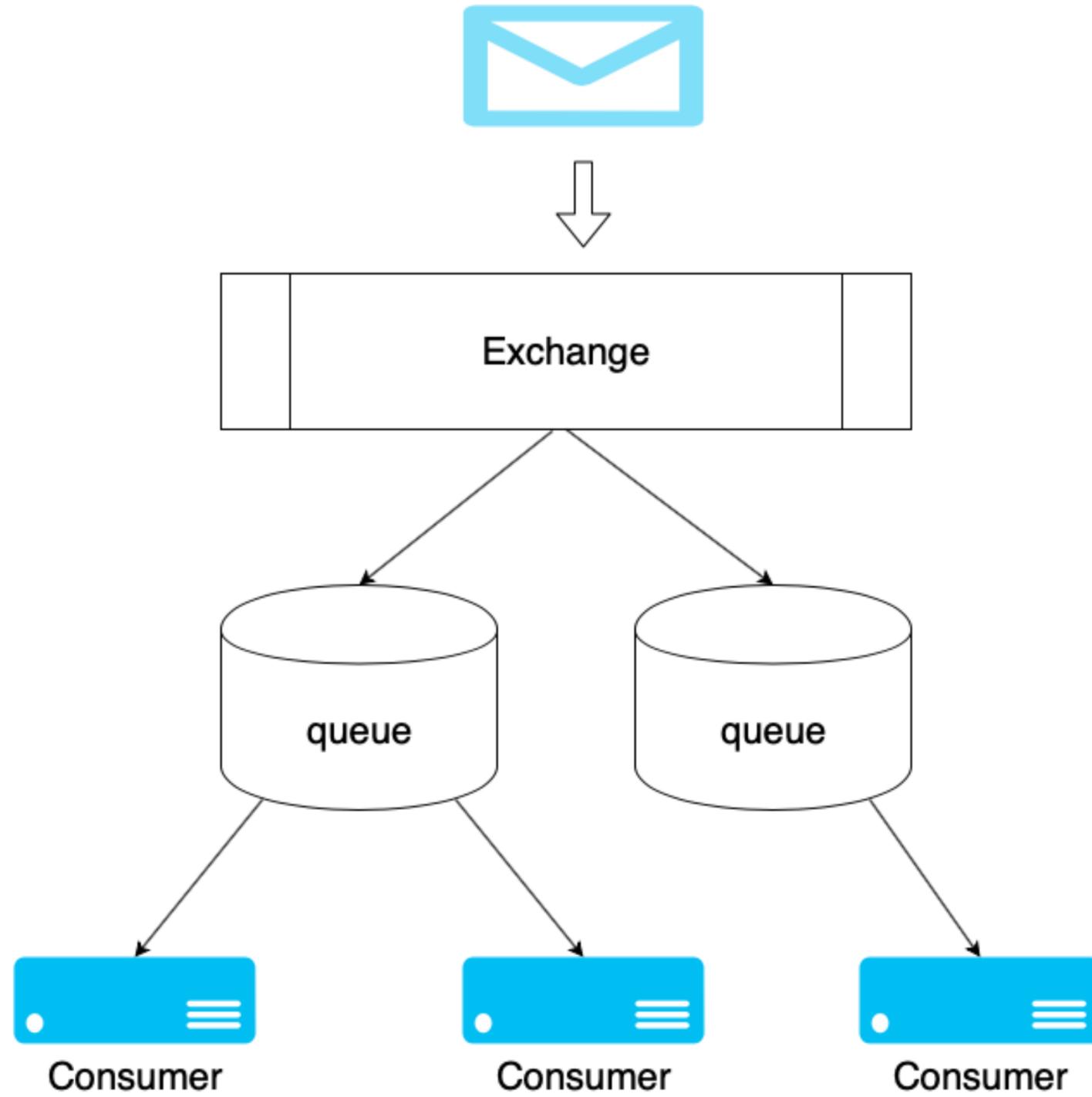
$3 * 10^6 - 5$  GB/M

$30 * 10^6 - 4$  GB/M

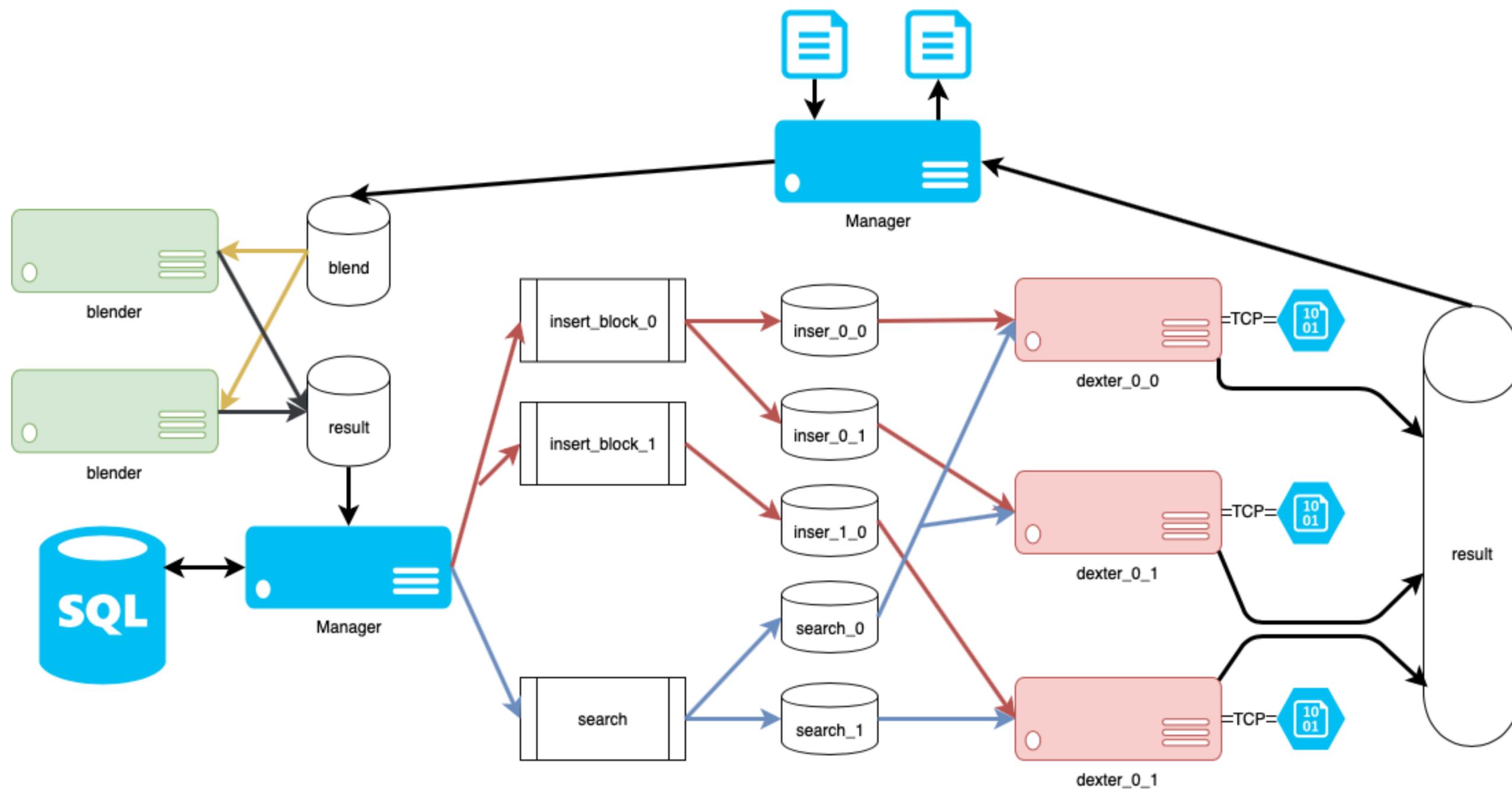


# RabbitMQ

---



# Конвейер поиска

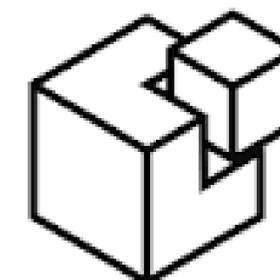


# Разработка и администрирование

---



PostgreSQL



SALTSTACK



REACTPHP

**Nagios**<sup>®</sup>



VAGRANT



CACTI

# Заключение

---



- Для rocket science нужно только желание
- Знания из ВУЗа лишними не бывают
- Pet проекты развивают
- Выступления помогают копнуть вглубь



# Вопросы?

**[MAIL@AVALLAC.RU](mailto:MAIL@AVALLAC.RU)**  
**<https://t.me/AVAIAC>**