

Одиннадцатая независимая научно-практическая конференция «Разработка ПО 2015»

22 - 24 октября, Москва



# Применение методов машинного обучения в системах хранения данных

Светлана Лазарева



# Задача на основе входящих запросов к СХД

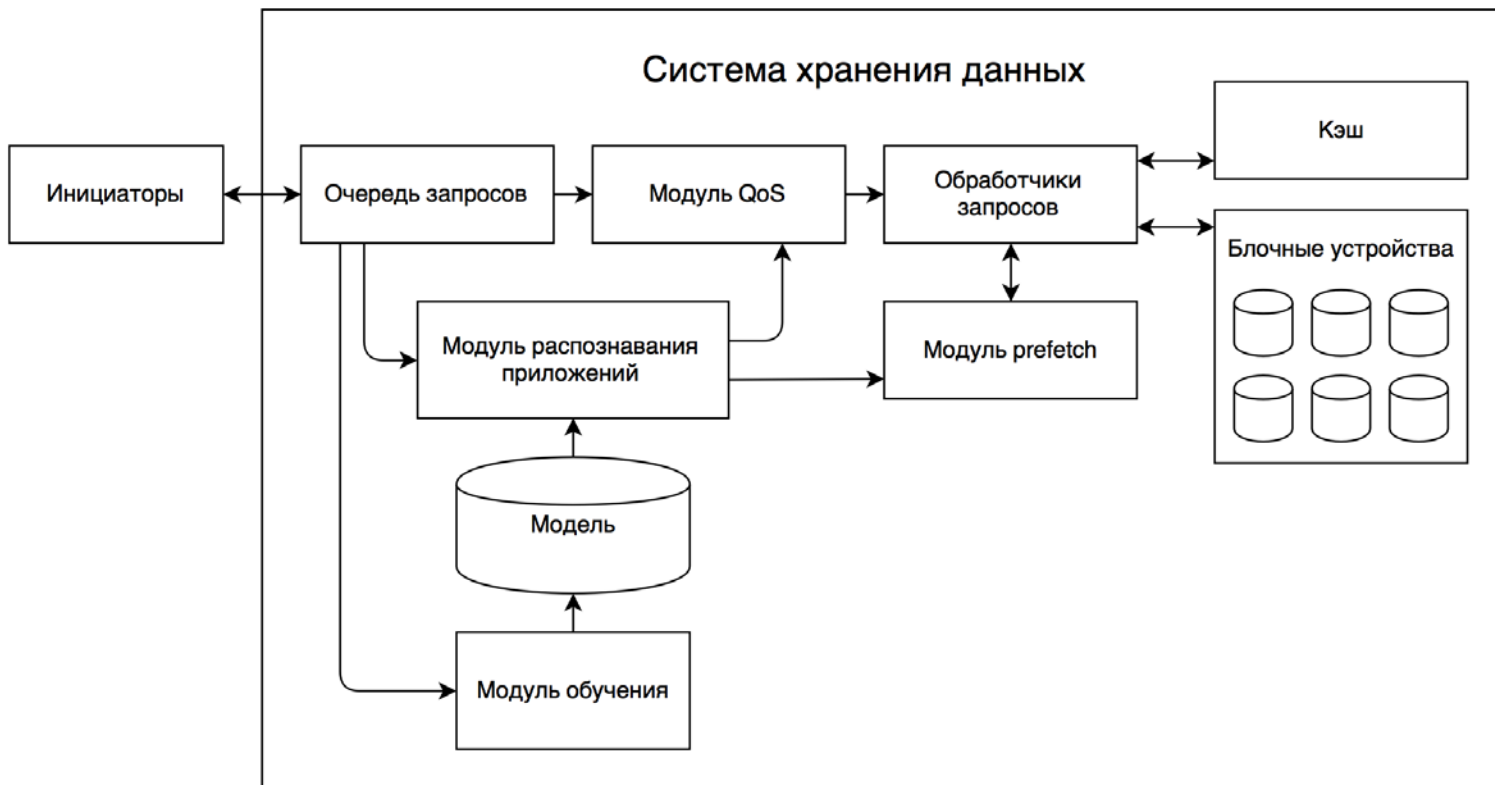
Автоматически выставлять  
приоритет приложениям



уметь предсказывать будущие  
запросы



**FUTURE  
PREDICTIONS**





# Требования к модулю распознавания



lightweight



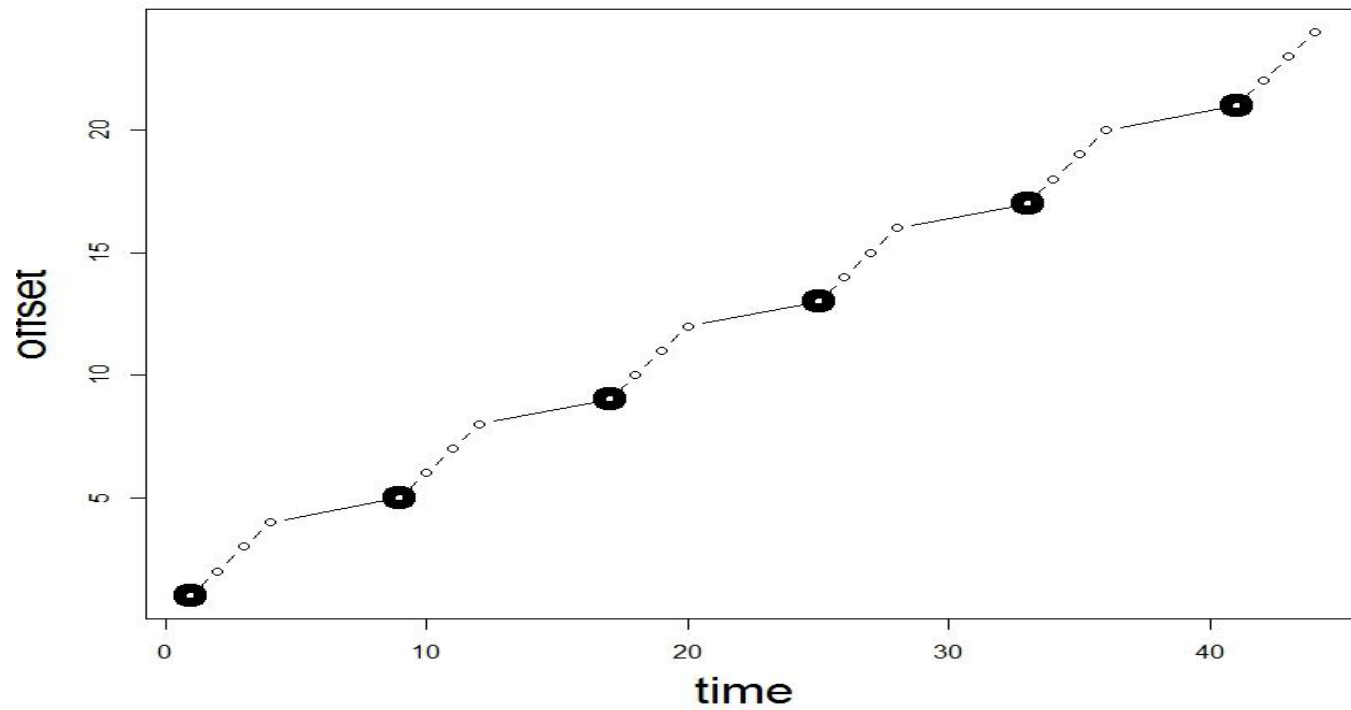
Accuracy



# Методология

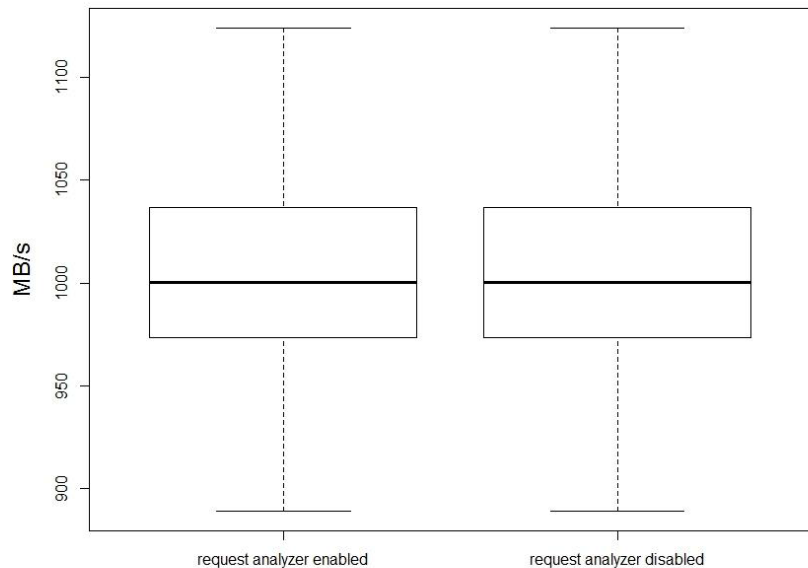
- Сбор трейсов от реальных клиентов
- Формирование сигнатур
- Классификатор- Random Forest
- Тестирование точности предсказания на реальных системах, тестирование на бенчмарках в условиях максимальной загрузки системы

## Pattern

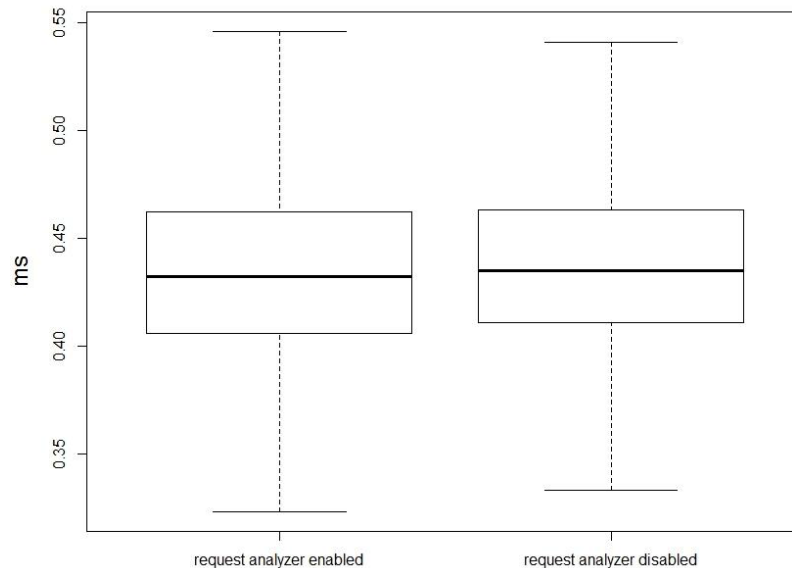


# FIO – максимальная загрузка

Boxplot of bandwidth distribution



Boxplot of latency distribution





# IOMETER

WORKLOAD	LENGTH OF REQUEST, KB	READ %	BW WITH ANALYZER/BW WITHOUT ANALYZER
Sequential read	4,32,256,1024	100%	1.3%
Sequential write	4,32,256,1024	0%	<1%
Sequential read/write	4,32,256,1024	50%	1.5%
Random read	4,32,256,1024	100%	1.4%
Random write	4,32,256,1024	0%	< 1%
Random read/write	4,32,256,1024	50%	2 %



## Точность предсказания. Реальные системы

	ERROR OF FIRST TYPE	ERROR IN SECOND TYPE
Apple Final Cut Pro/X	0.1%	0.5%
Adobe Premiere Pro	0.15%	0.8%
Autodesk Smoke	0.12%	0.7%
Antivirus Kaspersky Small Business Security	0.01%	0.01%
VMWare application server operating on MSSQL database	0.005%	0.01%
Non-priority class	0.1%	--



Спасибо за внимание!

[www.raidix.ru](http://www.raidix.ru)  
[request@raidix.com](mailto:request@raidix.com)