



# СРЕДА ЗА СЪХРАНЕНИЕ, ИЗВЛИЧАНЕ И КУРИРАНЕ НА ДАННИ ОТ ОБЛАСТТА НА ХУМАНИТАРИСТИКАТА И СОЦИАЛНИТЕ НАУКИ. ТЕХНОЛОГИЧНИ АСПЕКТИ

---

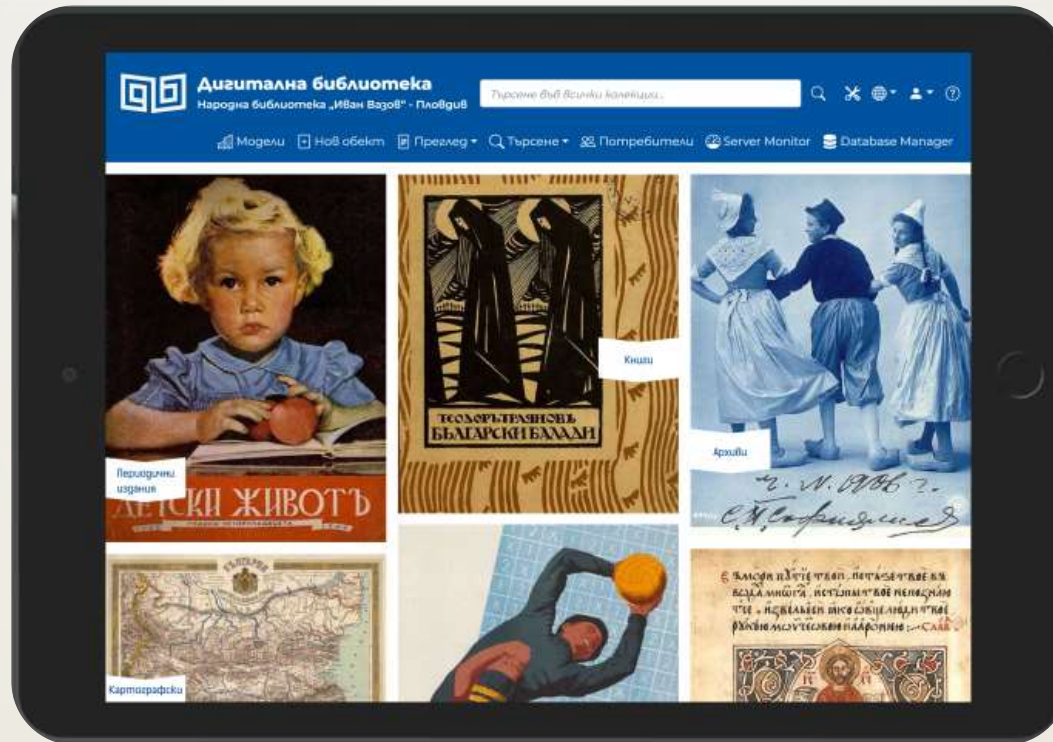
ДЕН НА ОТВОРЕНИТЕ ВРАТИ НА ЕЛЕКТРОННАТА ИНФРАСТРУКТУРА ЗА БЪЛГАРСКОТО  
ЕЗИКОВО И КУЛТУРНО НАСЛЕДСТВО КЛАДА-БГ, 16.11.2022, СОФИЯ, БЪЛГАРИЯ

Докладчик: Максим Гойнов

# СРЕДА ЗА СЪХРАНЕНИЕ, ИЗВЛИЧАНЕ И КУРИРАНЕ НА ДАННИ ОТ ОБЛАСТТА НА ХУМАНИТАРИСТИКАТА И СОЦИАЛНИТЕ НАУКИ



<https://bidl.math.bas.bg/>

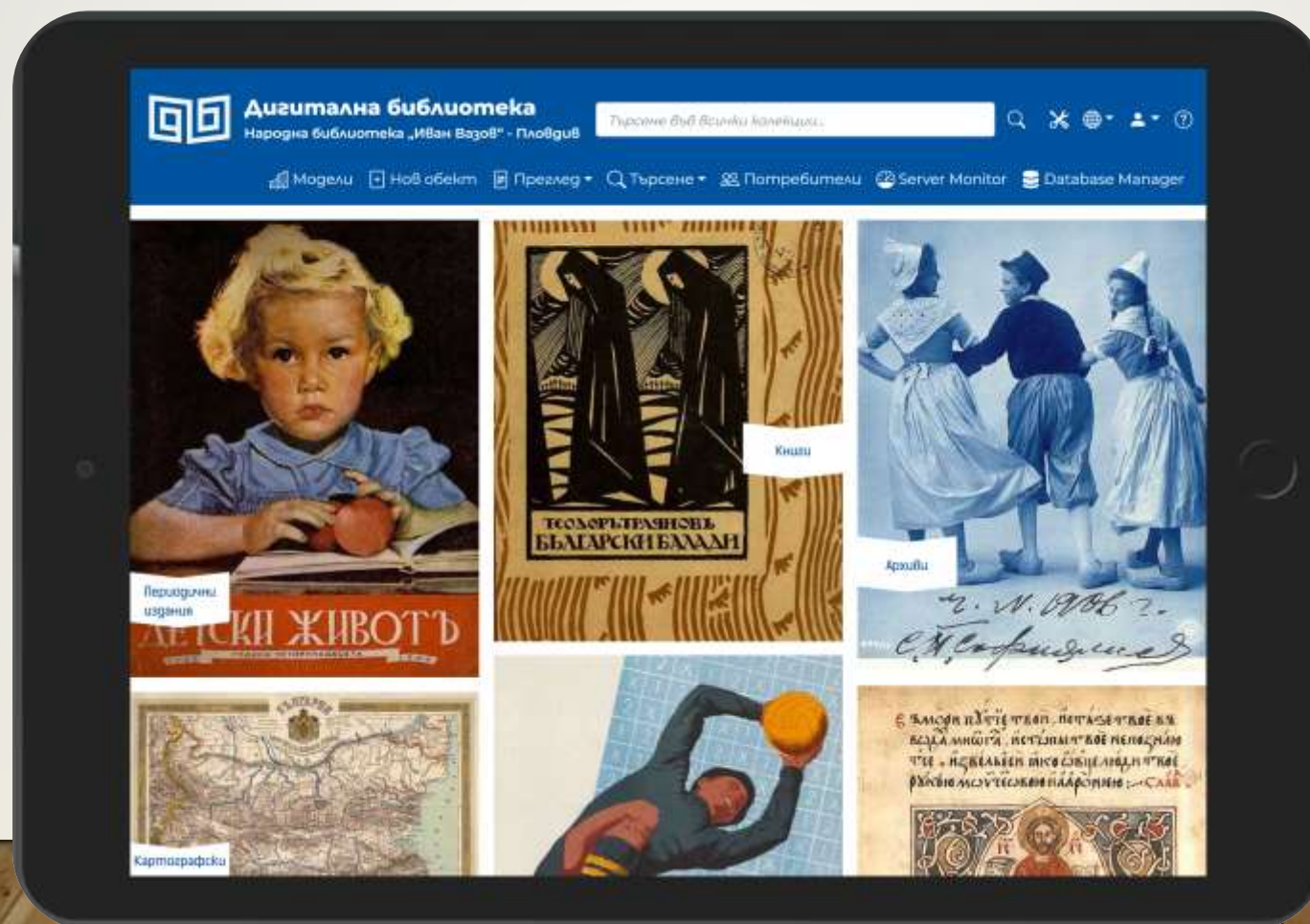


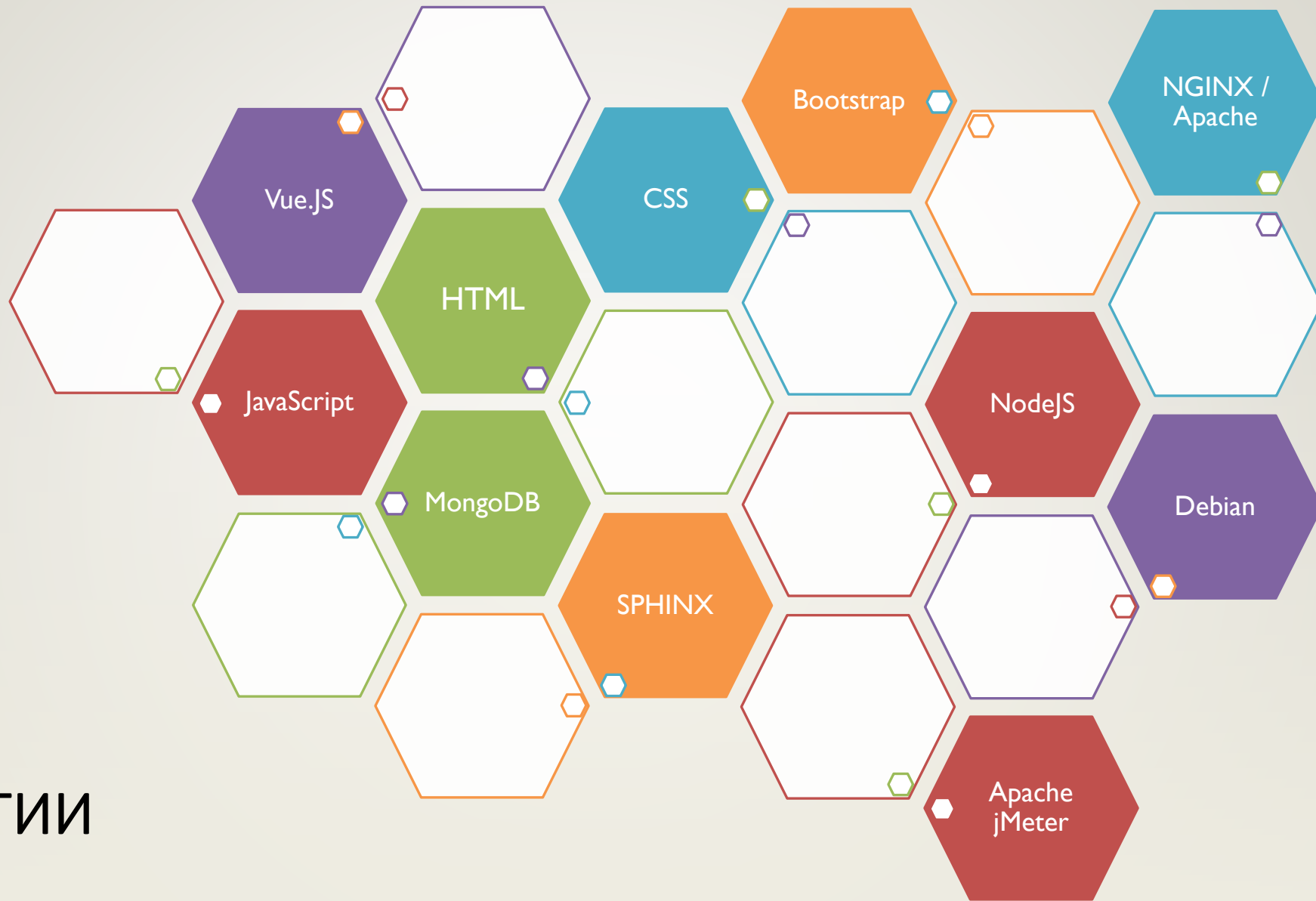
<https://digital.libplovdiv.com/>



<https://plakati.bg73.net/>

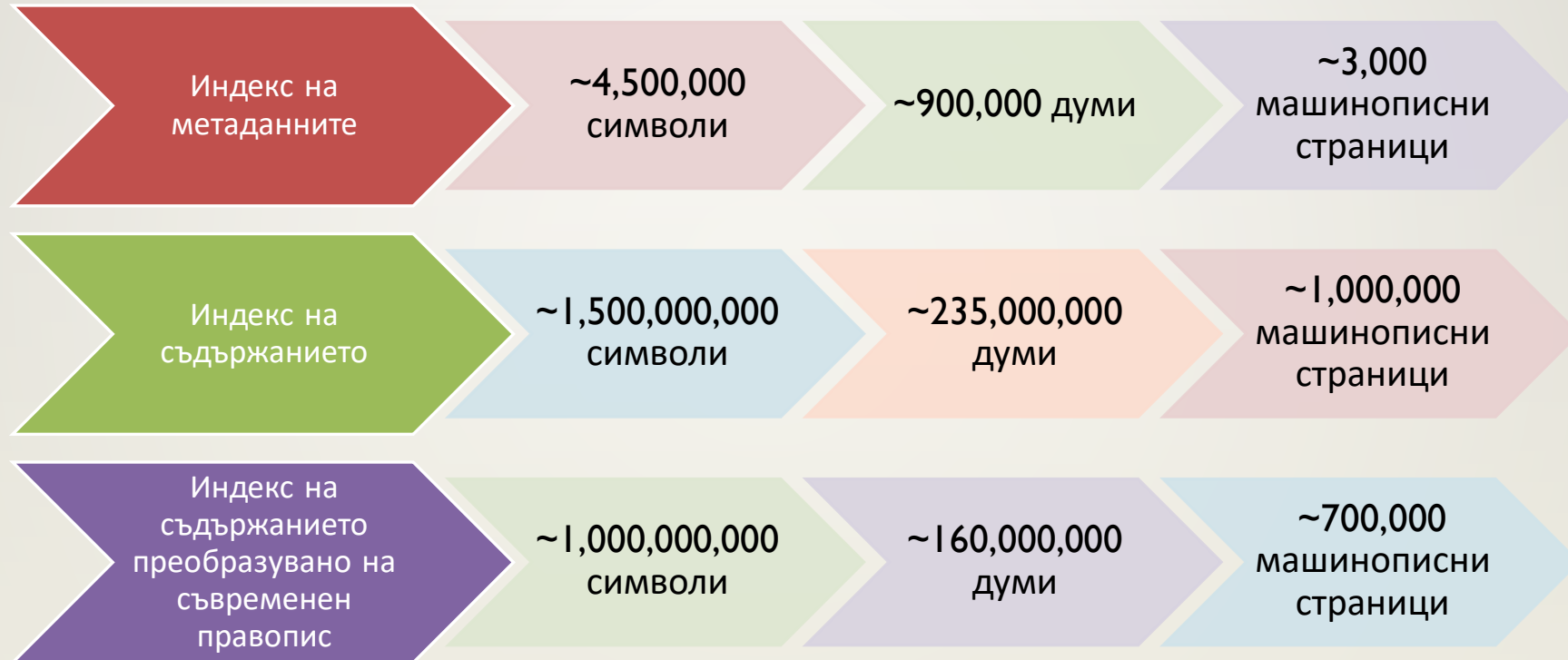
# ДИГИТАЛНА БИБЛИОТЕКА НАРОДНА БИБЛИОТЕКА „ИВАН ВАЗОВ“ – ПЛОВДИВ





ТЕХНОЛОГИИ

# ПЪЛНОТЕКСТОВО ТЪРСЕНЕ



# ПЪЛНОТЕКСТОВО ТЪРСЕНЕ

## MongoDB Community Server

- ~10 секунди за сложно търсене (повече от 1 дума в комбинация)
- Сериозно натоварване на CPU при изпълнение на заявката

## ElasticSearch

- <1 секунда за сложно търсене
- >10 секунди прехвърляне на резултата (при >10K резултати) към основната БД
- Ангажиране на голямо количество RAM памет

## SPHINX Search

- <1 секунда за сложно търсене
- <1 секунда за прехвърляне към БД
- Минимум RAM & CPU

# ПРОИЗВОДИТЕЛНОСТ



Apache

- 99.9% успешни заявки

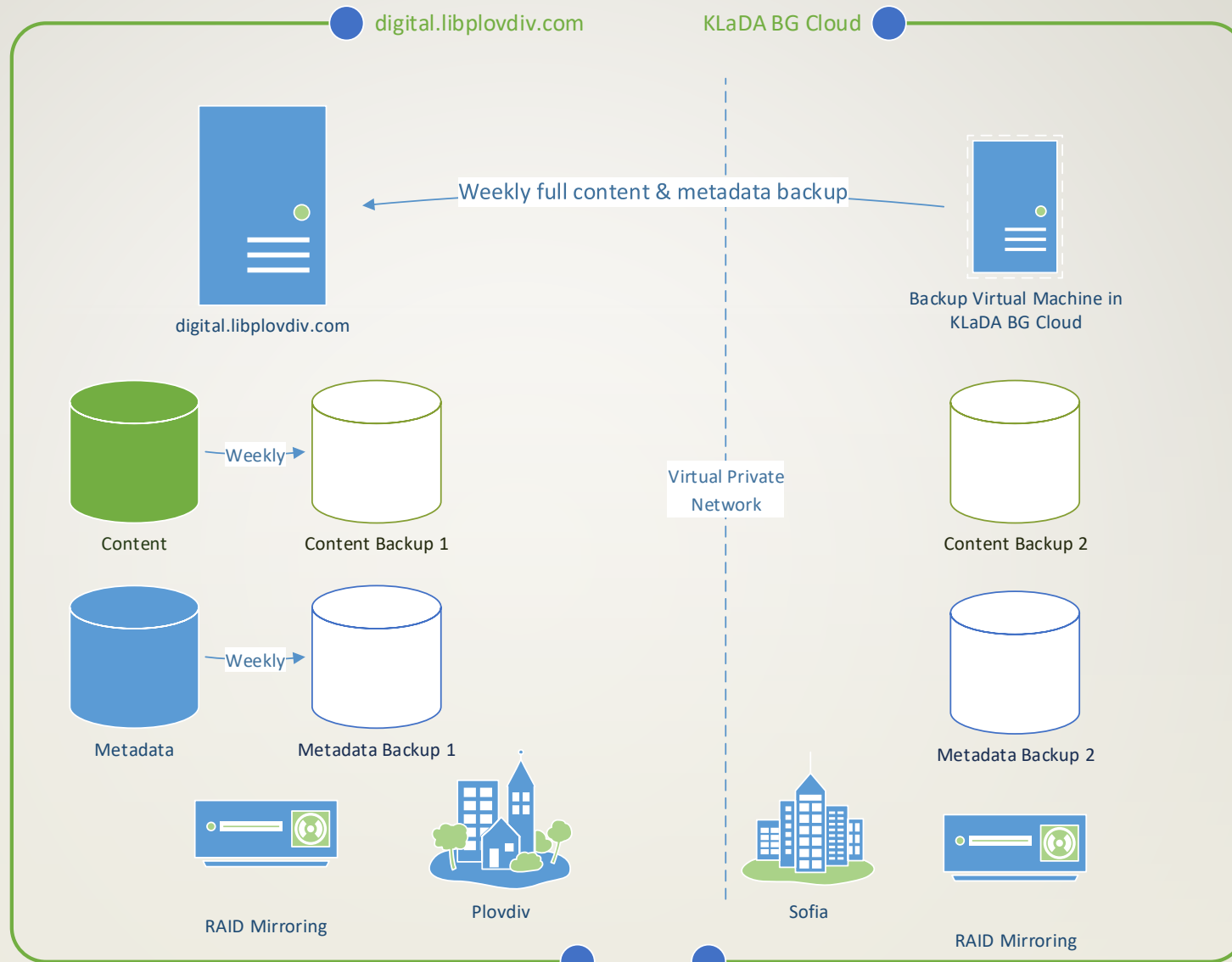
NGINX

- 100% успешни заявки



	Общ тест	Пълнотекстово търсене - тежки заявки
Брой заявки	129000	5750
Продължителност	0:59	4:50
Време за отговор (ms)		
Средно	260.67	2444.15
Медиана	202.56	2352.92
90% процентил	385.64	3161.38
95% процентил	493.79	3422.31
99% процентил	1500.94	4144.38
Минимално	3.00	9.00
Максимално	18240.00	6148.00
Средна стойност на RAM памет	4.6GB	5.3GB
Средна натовареност на процесорите	30.00%	92.00%
% неуспешни заявки	0.00	0.00

# СИГУРНОСТ. СЪХРАНЕНИЕ НА ДАННИТЕ





# СИГУРНОСТ. ПОТРЕБИТЕЛСКИ РОЛИ ИСТОРИЯ НА ПРОМЕНИТЕ

### Права

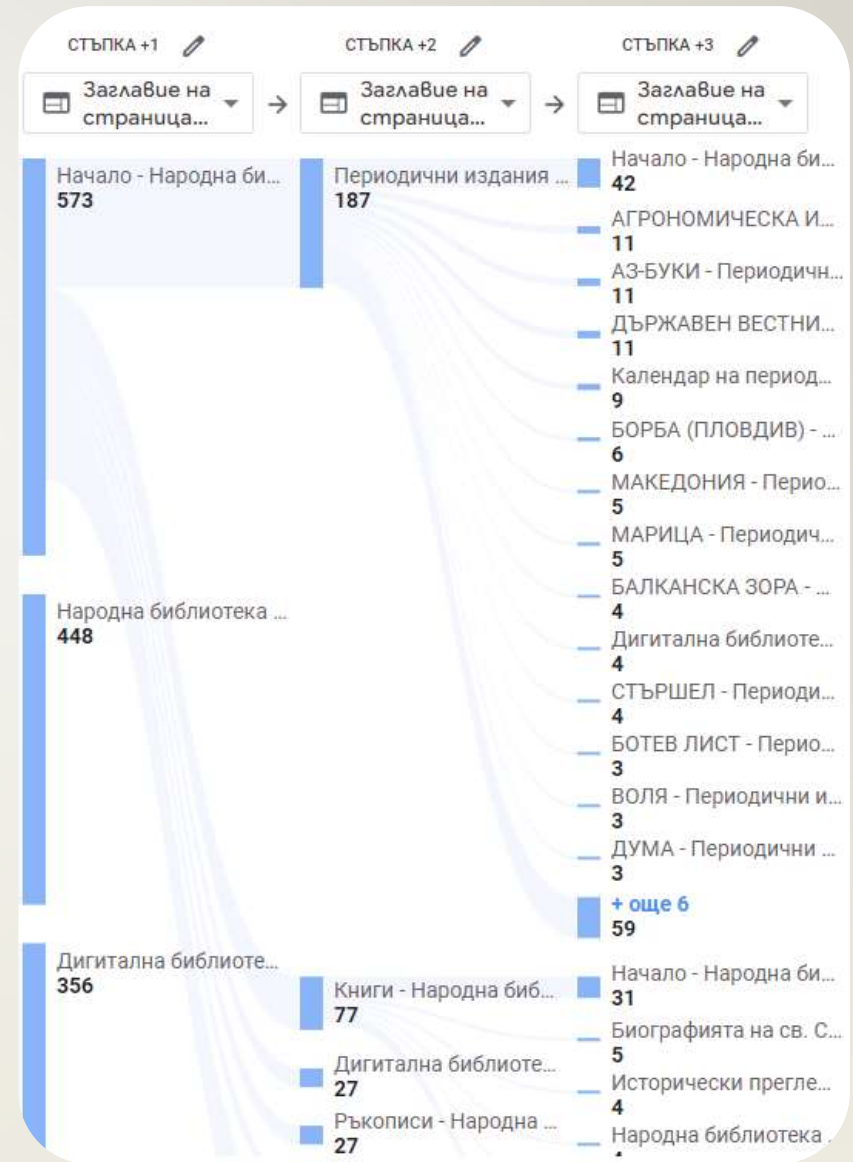
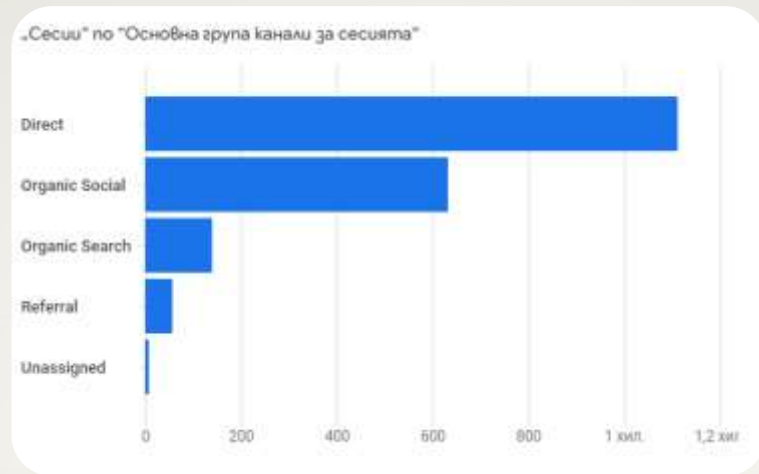
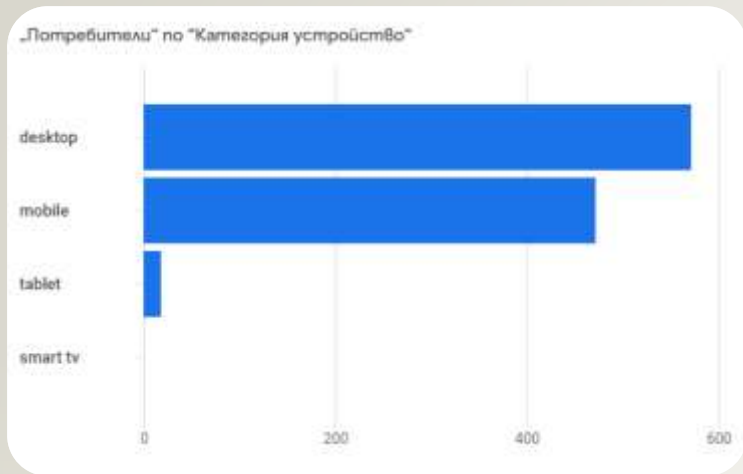
	Създаване	Четене	Запис	Изтривяне	Четене (създаден)	Запис (създаден)	Изтривяне (създаден)
<input type="checkbox"/> Администратор на БД	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Редактор на модели	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Администратор	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Редактор	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Потребител	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Гост	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Групи

Group rules

### История

Id	Action	Time	User	Type
	update	10/5/2022, 9:27:15 AM	harkins@gmail.com	current
	update	9/21/2022, 3:24:51 PM	g@r@n@l.com	history
	update	9/20/2022, 9:31:23 PM	g@r@n@l.com	history
	update	9/20/2022, 9:30:36 PM	g@r@n@l.com	history
	update	9/19/2022, 3:44:22 PM	g@r@n@l.com	history
	update	6/13/2022, 4:13:28 PM	g@r@n@l.com	history
	update	8/19/2021, 10:45:10 PM	g@r@n@l.com	history
	update	8/17/2021, 11:26:23 AM	g@r@n@l.com	history
	update	7/30/2021, 9:17:42 PM	g@r@n@l.com	history
	create	7/30/2021, 9:14:02 PM	g@r@n@l.com	history



# СТАТИСТИКИ – GOOGLE ANALYTICS



# МОНИТОРИНГ - NETDATA

## System Overview

Overview of the key system metrics.



## cpu

Total CPU utilization (all cores), 100% here means there is no CPU idle time at all. You can get per core usage at the [CPUs](#) section and per application usage at the [Applications Monitoring](#) section. Keep an eye on **lowait** ( 3.04%). If it is constantly high, your disks are a bottleneck and they slow your system down.

An important metric worth monitoring, is **softirq** ( 0.0133%). A constantly high percentage of softirq may indicate network driver issues.



Pressure Stall Information identifies and quantifies the disruptions caused by resource contentions. The "some" line indicates the share of time in which at least **some** tasks are stalled on CPU. The ratios (in %) are tracked as recent trends over 10-, 60-, and 300-second windows.

