

**«Открытый квотировано-кластерный
оркестратор ИИ (ОКОИИ) на базе
архитектуры PBDR "OQOAI:PBDR 5.2"»
Руководство пользователя**

Работа с интерфейсом ПО

Панели интерфейса

Стартовый интерфейс

Интерфейс до входа состоит из 3-х основных панелей

Варианты наполнения панелей контентом зависят от используемого режима

The screenshot shows the 'OQOAI:PBDR Панель администрирования' interface. At the top, it indicates 'Система онлайн' and 'Last update: 06:58:29'. The main section is divided into three panels:

- Panel 1 (Statistics and Management):** Displays 'Всего устройств: 3', 'Онлайн: 0', 'Офлайн: 0', and 'Группы: 1'. It includes buttons for 'Сканирование сети', 'Обновить статус', 'Добавить устройство', 'Добавить группу', and 'Метрики ОРУ'. Below these are tabs for 'Все узлы', 'Клиенты', and 'Сервера'.
- Panel 2 (Node List):** A table with columns: 'Имя узла', 'IP', 'Тип', 'Статус', 'Порт', 'Метрики', and 'Действия'. It lists three nodes: 'Scan_Server-02', 'Scan_Client-02', and 'Scan_Server-01'.
- Panel 3 (Groups):** A table with columns: 'Имя группы', 'Описание', 'Устройства', and 'Действия'. It lists one group: 'Продакт'.

1. Панель статистики и управления
2. Панель узлов
3. Панель групп

Верхняя панель статистики

Верхняя панель статистики Всего устройств отображает:

Общее количество добавленных узлов

Онлайн — количество доступных узлов

Офлайн — количество недоступных узлов

Группы — количество созданных групп

This is a close-up of the top statistics panel, showing four white boxes with black text and colored numbers:

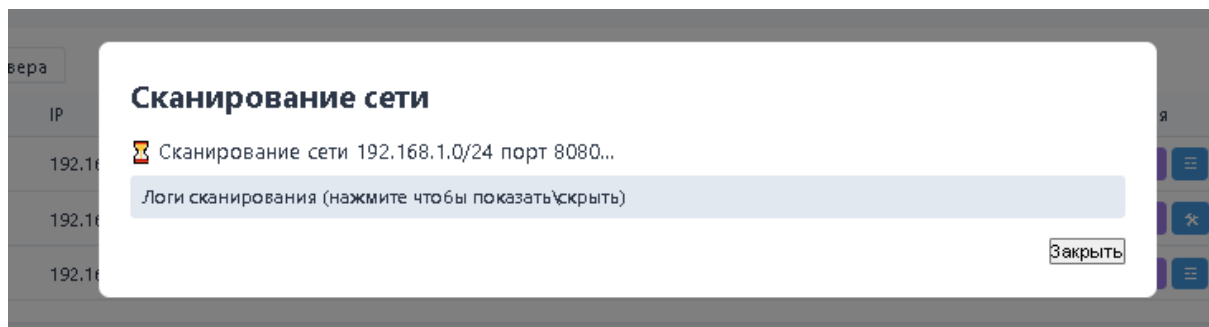
- Всего устройств: 3
- Онлайн: 0
- Офлайн: 0
- Группы: 1

Кнопки управления

Кнопка	Назначение
Сканирование сети	Поиск устройств в заданном диапазоне IP
Обновить статус	Принудительная проверка доступности всех узлов
Добавить устройство	Ручное добавление узла
Добавить группу	Создание группы устройств
Метрики GPU	Просмотр метрик всех серверов

Сканирование сети

1. Укажите **диапазон IP** (по умолчанию **192.168.1.0/24**)
2. Укажите **порт** (по умолчанию **8080**)
3. Нажмите **Сканирование сети**



Найденные устройства автоматически получают имена:

- Серверы: **Scan_Server-01**, **Scan_Server-02...**
- Клиенты: **Scan_Client-01**, **Scan_Client-02...**

Логи сканирования скрыты под спойлером -нажмите для просмотра.

Обновить статус

Принудительная проверка доступности всех узлов. Каждый узел опрашивается, после чего обновляется его статус (онлайн/офлайн).



Имя узла	IP	Тип	Статус	Порт	Метрики	Действия
Scan_Server-02	192.168.1.103	server	online	8082	-	   
Scan_Client-02	192.168.1.105	client	online	8080	-	   
Scan_Server-01	192.168.1.107	server	online	8080	-	   

Добавить устройство

Ручное добавление узла в список. Необходимо указать:

- **Имя** — произвольное имя узла
- **IP адрес** — сетевой адрес устройства
- **Тип** — сервер или клиент
- **Порт** — порт для подключения (по умолчанию 8080)



Добавить узел

Имя

IP адрес

Тип

Порт

Добавить группу

Создание группы устройств для массового применения конфигураций и политик.
Требуется указать:

- **Имя группы** — уникальное название
- **Описание** — краткое пояснение
- **Узлы** — список IP-адресов через запятую

Добавить группу

Имя группы

production-group

Описание

Production servers group

Узлы (IP, через запятую)

192.168.1.101, 192.168.1.102

Отмена

Добавить

Метрики GPU

Просмотр метрик всех серверов в кластере. Отображает:

- Использование GPU (%)
- Температуру GPU (°C)
- Свободную видеопамять (VRAM)
- Загрузку CPU (%)
- Длину очереди задач
- Количество выполненных заданий
- Загруженную модель
- Список доступных моделей

Метрики сервера

Scan_Server-02 (192.168.1.103)

Использование GPU
22.2%

Температура
50.8°C

VRAM свободно
4.9 GB

CPU нагрузка
N/A

Длина очереди
0

Задач всего
0

Загруженная модель
ornith-1.0-35b-Q4_K_M.gguf

Доступные модели
ornith-1.0-35b-Q4_K_M.gguf

Scan_Server-01 (192.168.1.107)

Использование GPU
17.6%

Температура
56.2°C

VRAM свободно
5.2 GB

CPU нагрузка
28.6%

Длина очереди
0

Задач всего
0

Загруженная модель
llama3.2:1b

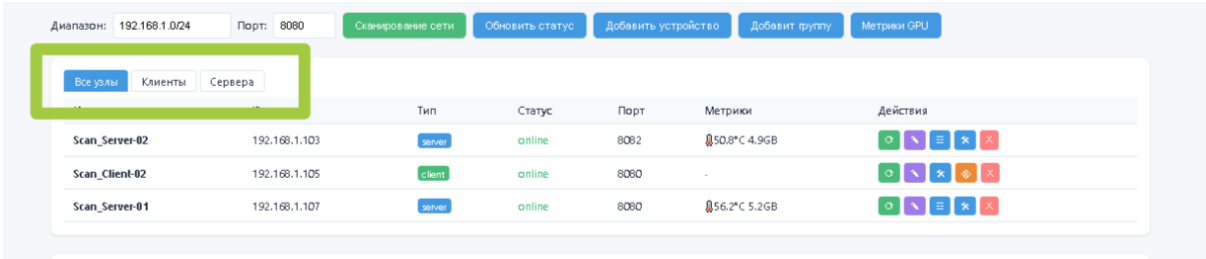
Доступные модели
llama3.2:1b

Заккрыть

Панель узлов

Вкладки узлов

- **Все узлы** — полный список устройств
- **Клиенты** — только клиенты
- **Сервера** — только серверы



Действия с узлами

Каждый узел в таблице имеет набор кнопок для управления:

Кнопка	Название	Назначение
	Обновить статус	Принудительная проверка доступности конкретного узла
	Редактировать	Изменение имени, IP-адреса, типа или порта узла
	Метрики	Просмотр детальных метрик сервера (только для серверов)
	Конфигурация	Загрузка и редактирование JSON-конфигурации узла
	Политики	Управление политиками маршрутизации (только для клиентов)
	Удалить	Удаление узла из списка



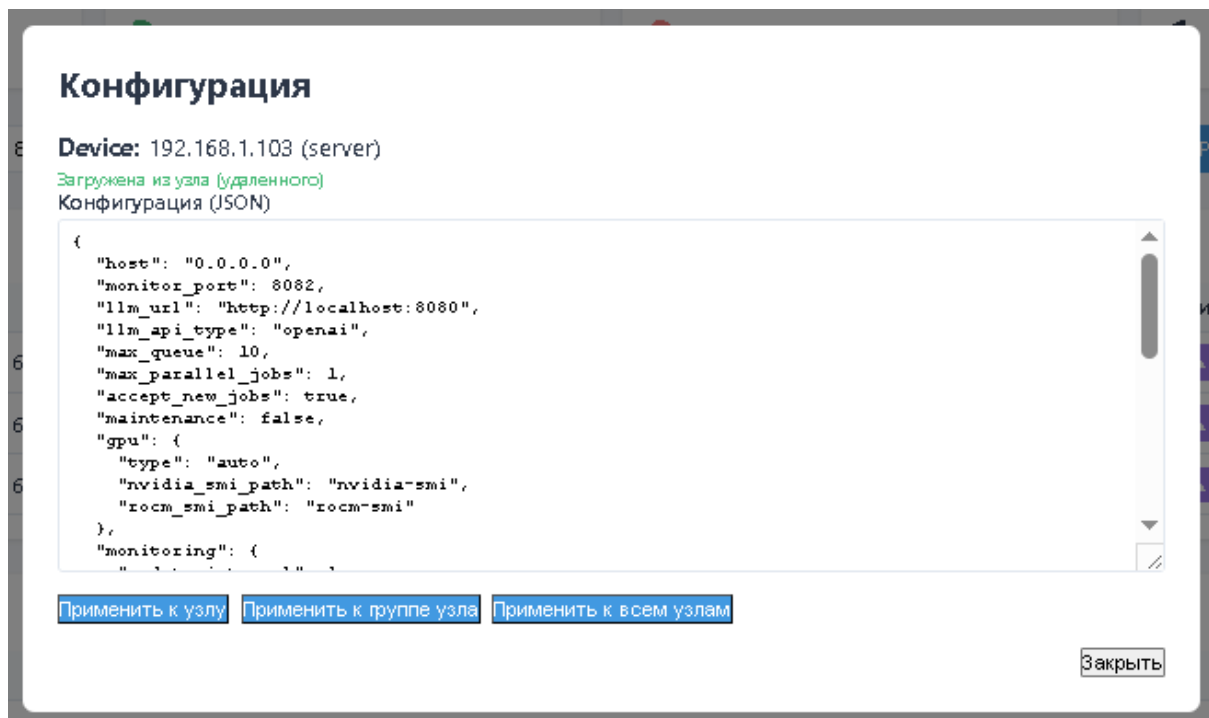
Конфигурация узла

При нажатии  открывается редактор конфигурации:

1. **Загружается JSON** с самого устройства (если узел доступен)
2. **Редактируйте** конфигурацию в текстовом поле
3. **Примените** изменения одной из кнопок:
 - **Применить к узлу** — отправить конфигурацию только на текущее устройство
 - **Применить к группе узла** — отправить на все устройства в группе (будет запрошено имя группы)
 - **Применить к всем узлам** — отправить на все устройства в списке

Индикатор источника конфигурации:

-  **Загружена из узла (удаленного)** — конфигурация успешно получена с устройства
-  **Загружена из кэша (узел не вернул конфигурацию)** — устройство недоступно, показана локальная копия



Политики маршрутизации (только для клиентов)

При нажатии  открывается редактор политик:

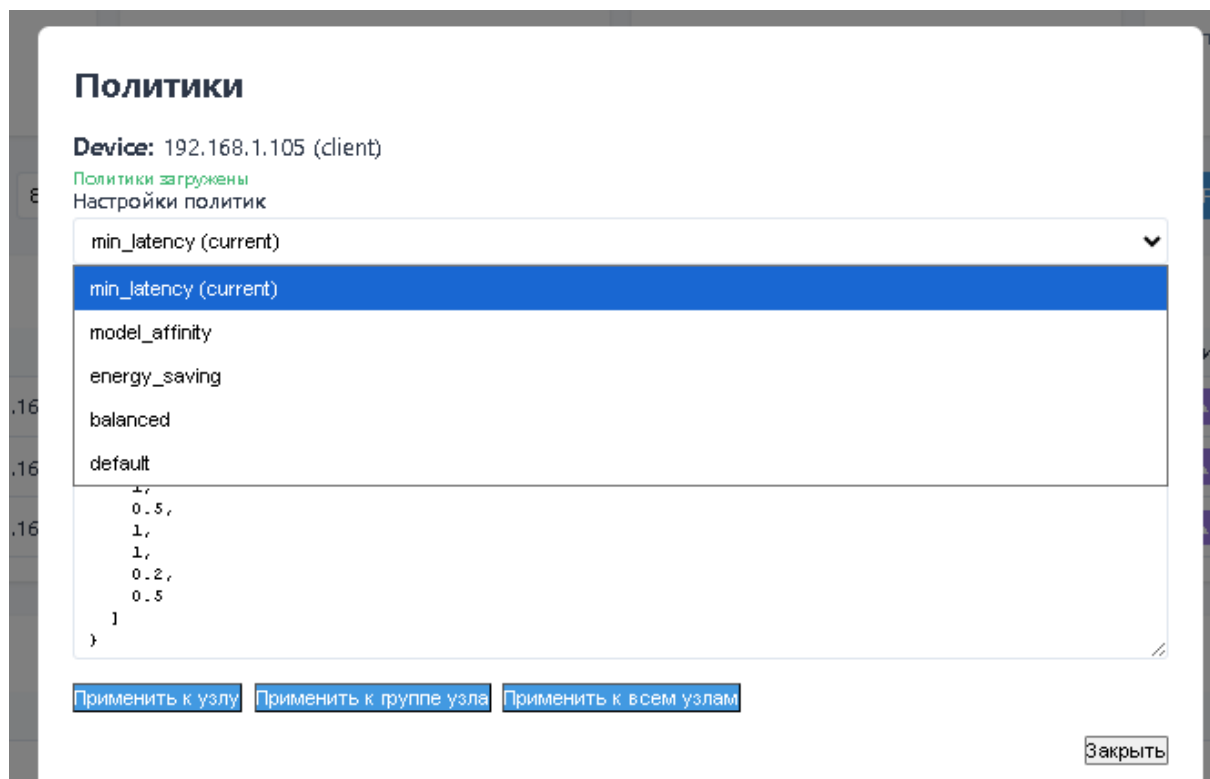
1. **Выберите политику** из выпадающего списка:

- **min_latency** — приоритет минимальной задержки
- **model_affinity** — приоритет узлов с уже загруженной моделью
- **energy_saving** — маршрутизация на наименее нагруженные узлы
- **balanced** — сбалансированный подход

2. **Посмотрите детали** политики в формате JSON (веса для каждого параметра)

3. **Примените** политику:

- **Применить к узлу** — сменить политику на текущем клиенте
- **Применить к группе узла** — сменить на всех клиентах в группе
- **Применить к всем узлам** — сменить на всех клиентах



Просмотр метрик сервера

При нажатии  на сервере открывается окно с детальными метриками:

Метрика	Описание
GPU нагрузка	Загрузка графического процессора в процентах
Температура	Температура GPU в градусах Цельсия
VRAM свободно	Свободная видеопамять в гигабайтах

Метрика	Описание
CPU нагрузка	Загрузка центрального процессора в процентах
Очередь задач	Количество задач в очереди
Задач всего	Общее количество выполненных задач
Загруженная модель	Текущая загруженная модель
Доступные модели	Список всех доступных моделей

Метрики сервера

Scan_Server-02 (192.168.1.103)

Использование GPU
6.1%

Температура
61.1°C

VRAM свободно
3.8 GB

CPU нагрузка
N/A

Длина очереди
0

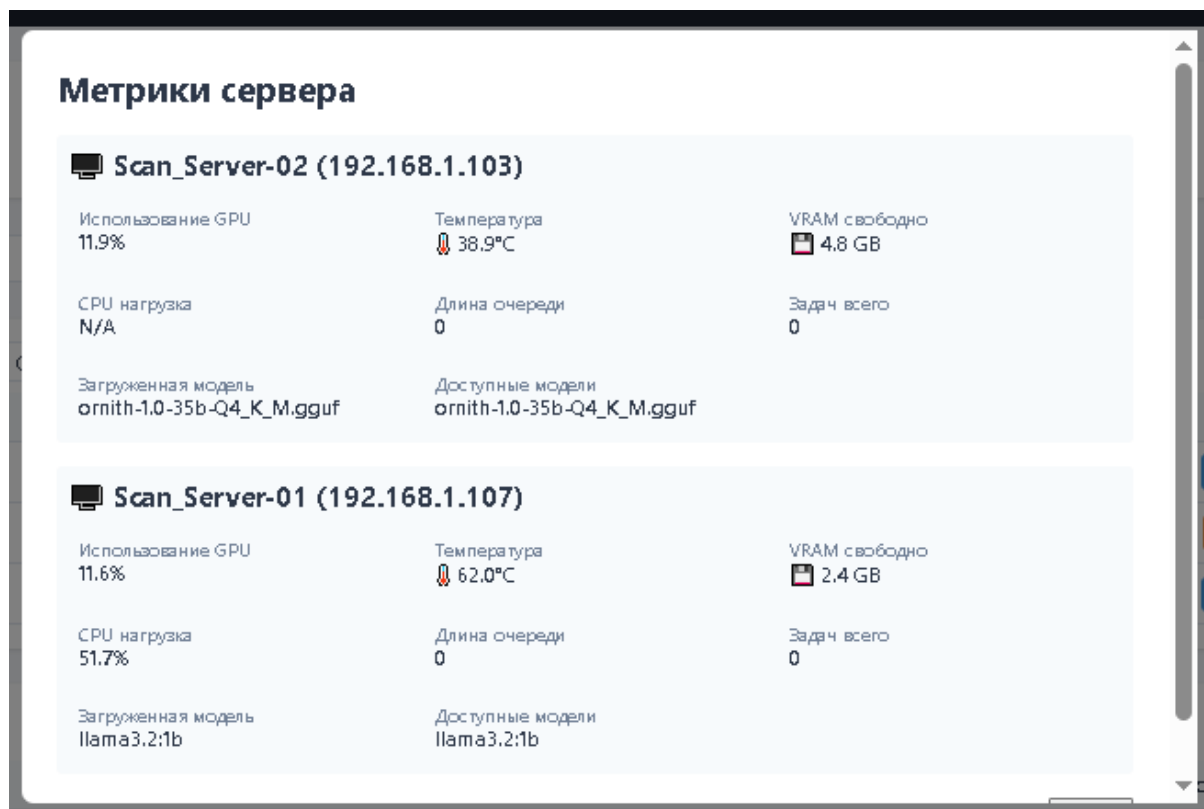
Задач всего
0

Загруженная модель
ornith-1.0-35b-Q4_K_M.gguf

ornith-1.0-35b-Q4_K_M.gguf

Закреть

Кнопка **Метрики GPU** на главной панели открывает метрики всех серверов одновременно.



Панель групп узлов

Группы позволяют массово применять конфигурации и политики.

Создание группы:

1. Нажмите **Добавить группу**
2. Укажите **имя группы** (уникальное)
3. Добавьте **описание** (опционально)
4. Перечислите **IP-адреса** через запятую: 192.168.1.101, 192.168.1.102

Добавить группу

Имя группы
production-group

Описание
Production servers group

Узлы (IP, через запятую)
192.168.1.101, 192.168.1.102

Отмена Добавить

Использование групп:

- При применении конфигурации или политики выберите "Применить к группе узла"
- Введите имя группы — изменения отправятся на все узлы в ней

Сохранение данных

Все изменения (добавление, удаление, редактирование узлов и групп) **автоматически сохраняются** в JSON-файл конфигурации админ-панели. Данные не теряются после перезапуска.

Формат конфигурации

Файлы конфигурации ПО хранят индивидуальные настройки и могут быть изменены из интерфейса администратора

Установка

Для установки ПО необходимо:

1. Разместить дистрибутив в папку пользователя,
2. Распаковать в папку PBDR
3. Выполнить запуск скрипта соответствующего типу узла (сервер, клиент, администратор), передав дополнительным атрибутом путь к файлу конфигурации

После запуска информация о текущем состоянии работы будет отображаться в консоли

```
082/status
2026-07-29 07:36:29.789 - PBDR-Client - INFO - PBDR Client <SYNC> running on http://0.0.0.0:8080
2026-07-29 07:36:29.791 - PBDR-Client - INFO - Debug: http://0.0.0.0:8080/debug
Updated: 192.168.1.107:11434
2026-07-29 07:36:29.901 - PBDR-Debug - DEBUG - Updated: 192.168.1.107:11434
Updated: 192.168.1.103:8080
2026-07-29 07:36:29.904 - PBDR-Debug - DEBUG - Updated: 192.168.1.103:8080
Discovery complete: 2 nodes
2026-07-29 07:36:29.906 - PBDR-Debug - DEBUG - Discovery complete: 2 nodes
```

```
2026-07-29 07:38:20.661 - PBDR-Server - INFO - Hardware Monitor initialized with GPU: none
2026-07-29 07:38:20.661 - PBDR-Server - INFO - PBDR Server initialized
2026-07-29 07:38:20.662 - PBDR-Server - INFO - PBDR Server API running on http://0.0.0.0:8080
2026-07-29 07:38:20.685 - PBDR-Server - INFO - Ollama API detected
2026-07-29 07:38:20.685 - PBDR-Server - INFO - Auto-detected LLM API type: ollama
2026-07-29 07:38:20.687 - PBDR-Server - INFO - Discovered Ollama models: ['llama3.2:1b']
2026-07-29 07:38:21.931 - aiohttp.access - INFO - 192.168.1.105 [29/Jul/2026:07:38:21 +0300] "GET /status HTTP/1.1" 200
1384 "-" "Python/3.7 aiohttp/3.8.6"
2026-07-29 07:38:23.071 - aiohttp.access - INFO - 192.168.1.105 [29/Jul/2026:07:38:22 +0300] "GET /status HTTP/1.1" 200
1385 "-" "Python/3.7 aiohttp/3.8.6"
```

```
2026-07-29 07:40:24.144 - PBDR-Admin - INFO - Device Manager initialized with 3 devices
2026-07-29 07:40:24.148 - PBDR-Admin - INFO - Admin Server running on http://0.0.0.0:8081
2026-07-29 07:40:30.230 - aiohttp.access - INFO - 127.0.0.1 [29/Jul/2026:07:40:30 +0300] "GET /api/stats HTTP/1.1" 200
42 "http://localhost:8081/" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/149.0.0.0 Safari/537.36"
2026-07-29 07:40:30.231 - aiohttp.access - INFO - 127.0.0.1 [29/Jul/2026:07:40:30 +0300] "GET /api/devices HTTP/1.1" 200
884 "http://localhost:8081/" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/149.0.0.0 Safari/537.36"
```

Настройка эндпоинта запросов

Запросы клиента

В конфигурации каждого клиента указан хост и порт OpenAI API совместимого эндпоинта например:

"host": "192.168.1.105", - IP хоста запросов

"port": 8080 - порт хоста запросов

указание вместо IP хоста 0.0.0.0 позволяет обрабатывать внешние запросы узла,
Для конфигурации эндпоинта запросов конечного пользователя укажите URL в формате:

http://<IP хоста>:<порт хоста>/v1

например:

http://192.168.1.105:8080/v1

Запросы сервера

В конфигурации каждого сервера указаны хосты и порты эндпоинтов LLM серверов например

"host": "192.168.1.105", - IP LLM сервера

"port": 11434 - порт LLM сервера

Для конфигурации эндпоинта запросов конечного пользователя укажите URL в формате рекомендованном вашим LLM сервером:

Конфигурация сервера поддерживает установку нескольких вариантов форматов данных, например:

"llm_api_type": "openai", - для OpenAI API совместимого эндпоинта

"llm_api_type": "ollama", - для Ollama API совместимого эндпоинта

"llm_api_type": "auto", - для автоопределения формата