



УТВЕРЖДАЮ

Генеральный директор
ООО «Цифровые сервисы»


А.Н. Ваулин
«01» августа 2024 г.


**Платформа корпоративного распределенного хранилища больших
данных. Монитор.
(ПБД. Монитор)**

Руководство пользователя

на 13 стр.

Разработчик

Руководитель отдела
ООО «Цифровые сервисы»


А.В. Гайдабура
«29» июля 2024 г.

Оглавление

1	Общее положение.....	3
1.1	Введение	3
1.2	Назначение Расширения.....	3
1.3	Условия применения	3
1.3.1	Требования к конфигурации рабочего места.....	3
1.3.2	Требования к квалификации Пользователей.....	3
2	Работа на ПБД. Монитор	4
2.1	Мониторинг ETL конвейеров (извлечение, преобразование и загрузка данных).....	5
2.2	Мониторинг отставания потребителя Apache Kafka.....	6
2.3	Мониторинг состояния кластера Greenplum.....	8
2.4	Функция контроля работоспособности.....	11
3	Действия пользователей при возникновении ошибок.....	13

1 Общее положение

1.1 Введение

Руководство пользователя содержит описание работы пользователя с «Платформой корпоративного распределенного хранилища больших данных. Монитор», сокращенное название «ПБД. Монитор» (далее по тексту – Расширение).

Руководство пользователя является одним из документов эксплуатационной документации. Данное Руководство содержит подробное описание работы с Расширением, пошаговые инструкции по выполнению различных реализуемых им задач.

1.2 Назначение Расширения

Расширение предназначено для отслеживания функционального состояния компонент «Платформы корпоративного распределенного хранилища больших данных (КХД) (далее - Платформы).

1.3 Условия применения

1.3.1 Требования к конфигурации рабочего места

В качестве рабочего места должен использоваться персональный компьютер.

На рабочих местах пользователей должна быть следующая конфигурация программного обеспечения:

- Операционная система: Linux, Windows 7-10 и выше, RedOS.
- Браузер: Яндекс Браузер версии 23 и выше, Microsoft Edge версии 42.1 и выше.

1.3.2 Требования к квалификации Пользователей

Для выполнения своих должностных обязанностей с использованием Расширения пользователи должны обладать соответствующей квалификацией и быть ознакомлены с функциональностью Платформы.

2 Работа на ПБД. Монитор

Визуальный контроль функционального состояния компонент Платформы реализован средствами Grafana с использованием информационных панелей.

Перед использованием необходимо пройти авторизацию в WEB интерфейсе по ссылке: `http://{ groups['grafana'][0] }:3000/login`, где: `{ groups['grafana'][0] }` - соответствующий хост, FQDN отличается в зависимости от контура (также, существует возможность воспользоваться уже развернутым расширением, перейдя по ссылке: <http://rnd-edu-namenode-01.khd.lan:3000/> на демонстрационный стенд).

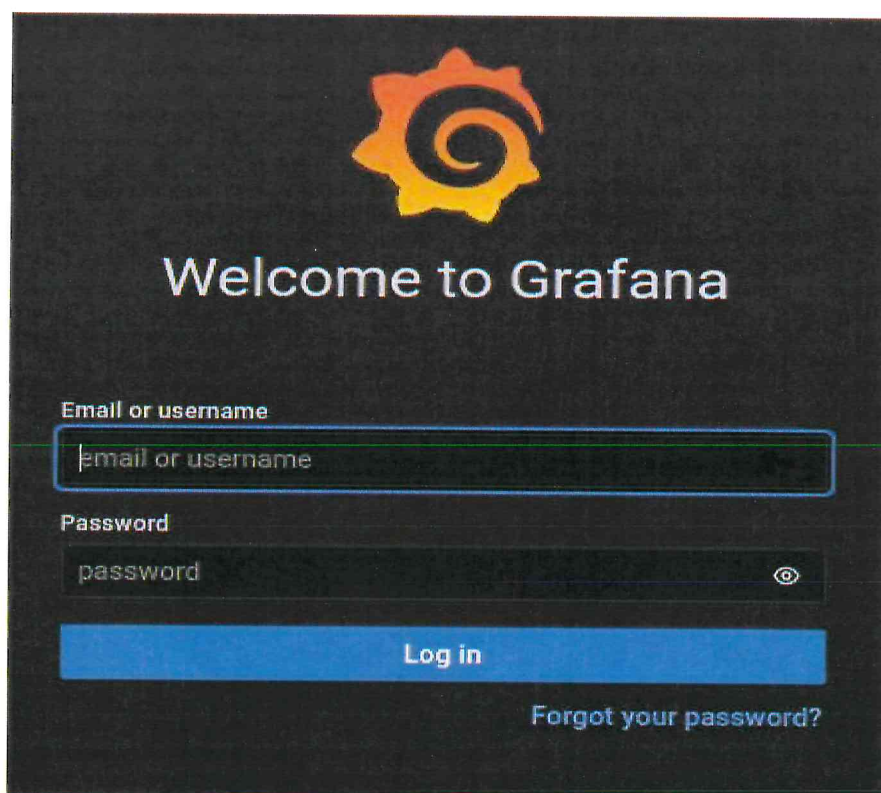


Рисунок 1. Окно авторизации

В поле «Имя пользователя» и «Пароль» ввести данные пользователя, обладающего правами администратора (изначальные политики доступа задаются с использованием административных учетных данных: логин «**admin**», пароль «**admin**»).

Нажать кнопку «Log in».

2.1 Мониторинг ETL конвейеров (извлечение, преобразование и загрузка данных)

Информационная панель для мониторинга использующихся ETL конвейеров, реализуемых как направленные ациклические графы (DAG), доступна в группе «General».

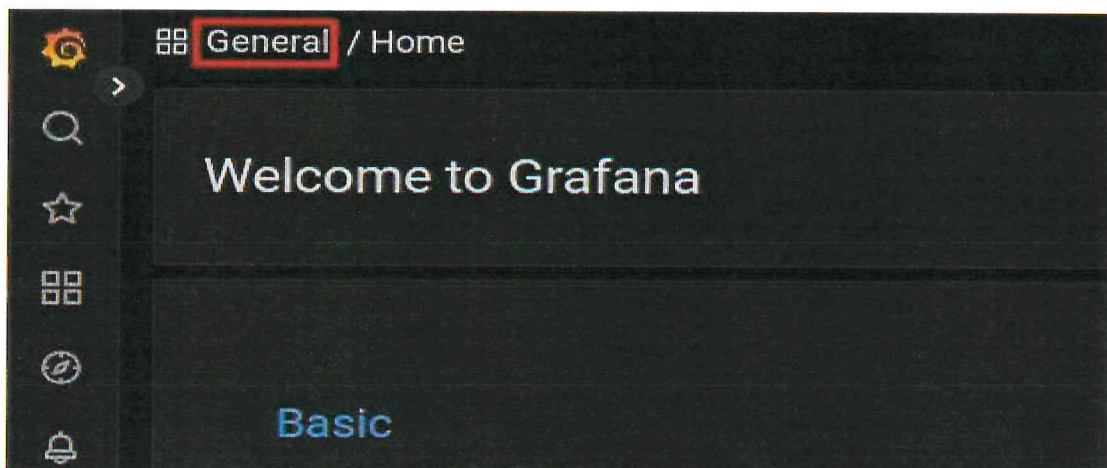


Рисунок 2. Расположение информационной панели

Для просмотра количества проваленных направленных ациклических графов (DAG) в разрезе owner+dag_id+task_id используется информационная панель «ETL monitoring».

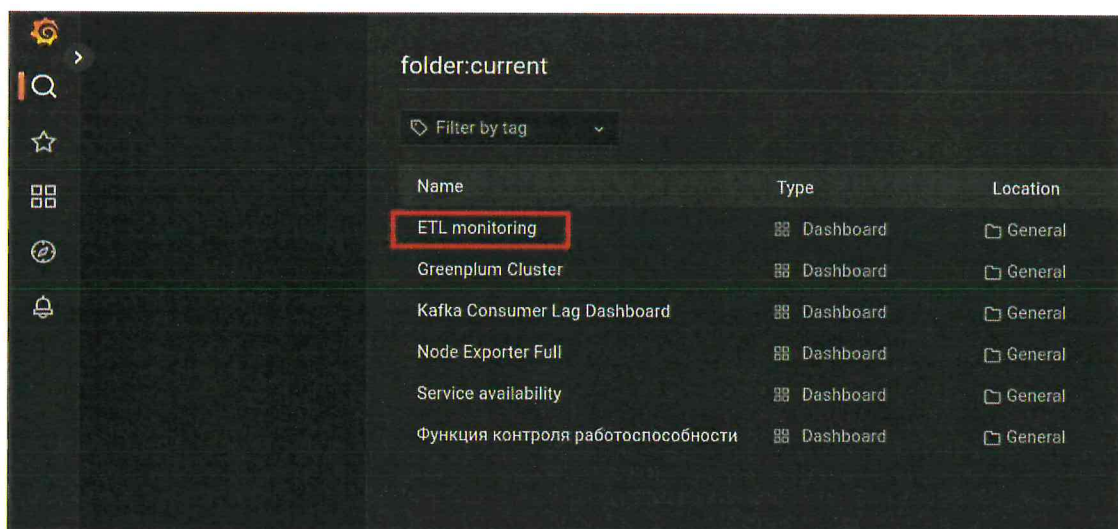


Рисунок 3. Мониторинг ETL конвейеров

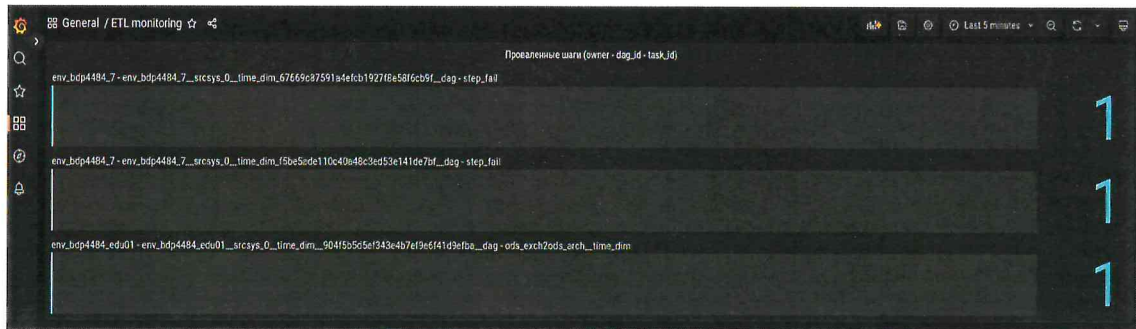


Рисунок 4. Визуальный контроль наличия проблем

2.2 Мониторинг отставания потребителя Apache Kafka

Информационная панель контроля отставания потребителя доступен в группе «General».

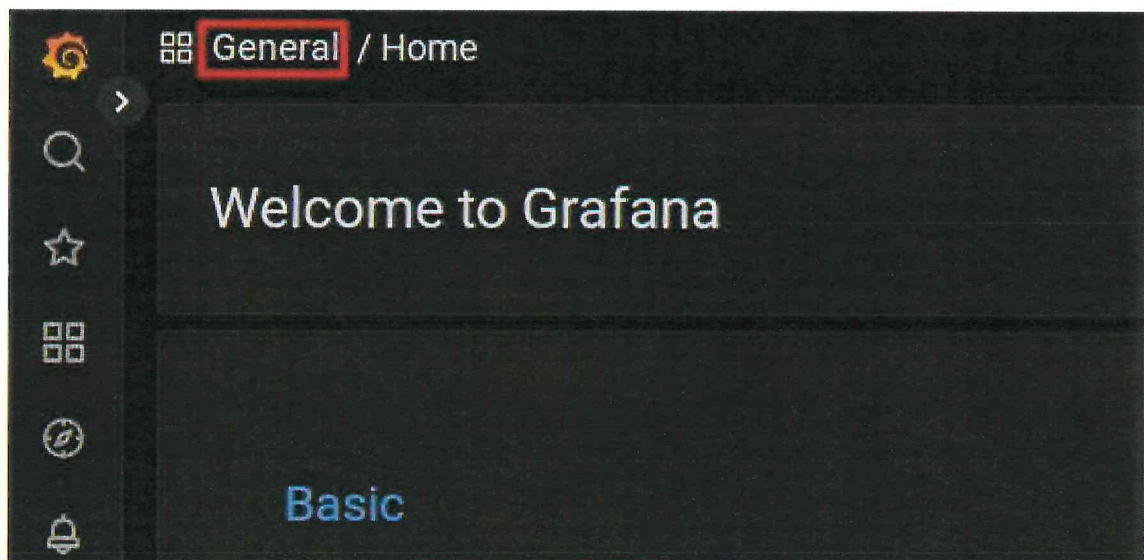


Рисунок 5. Расположение информационной панели

Для просмотра текущего состояния Kafka topics и контроля динамики изменений на временной оси используется информационная панель «Kafka Consumer Lag Dashboard».

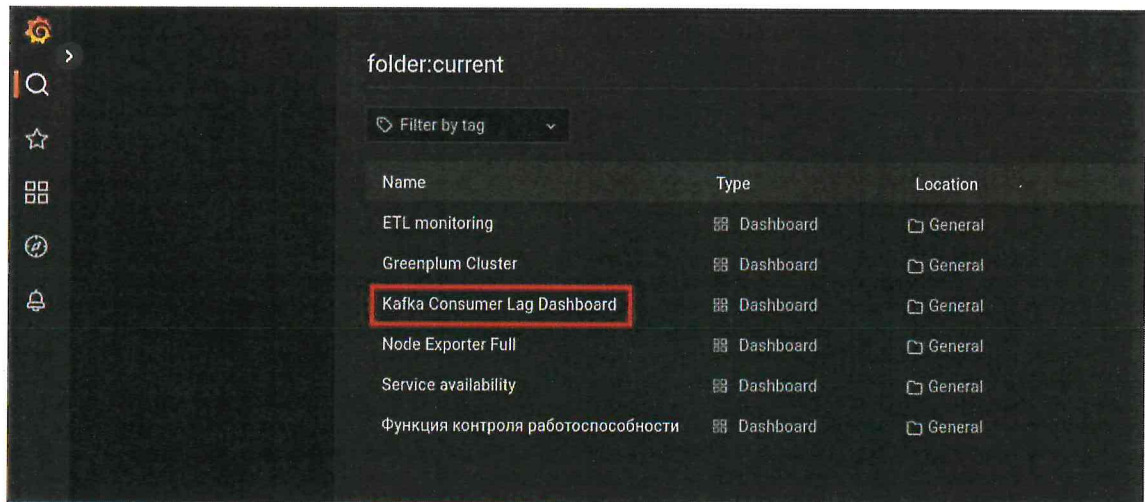


Рисунок 6. Мониторинг потребителя Apache Kafka

Инструмент, наглядно представляющий ситуацию - плитки «Consumer Group Lag Partition Offsets».



Рисунок 7. Визуальный контроль отставания потребителя Apache Kafka

Плитки «Consumer Group Lag Partition Offsets» - реализуются запросом получения топ-25 Kafka topics с наибольшей задержкой потребления.

Для просмотра максимального времени задержки группы потребителей в секундах и смещения разделов групп потребителей используются следующие плитки информационной панели:



Рисунок 8. Отставание всех групп потребителей

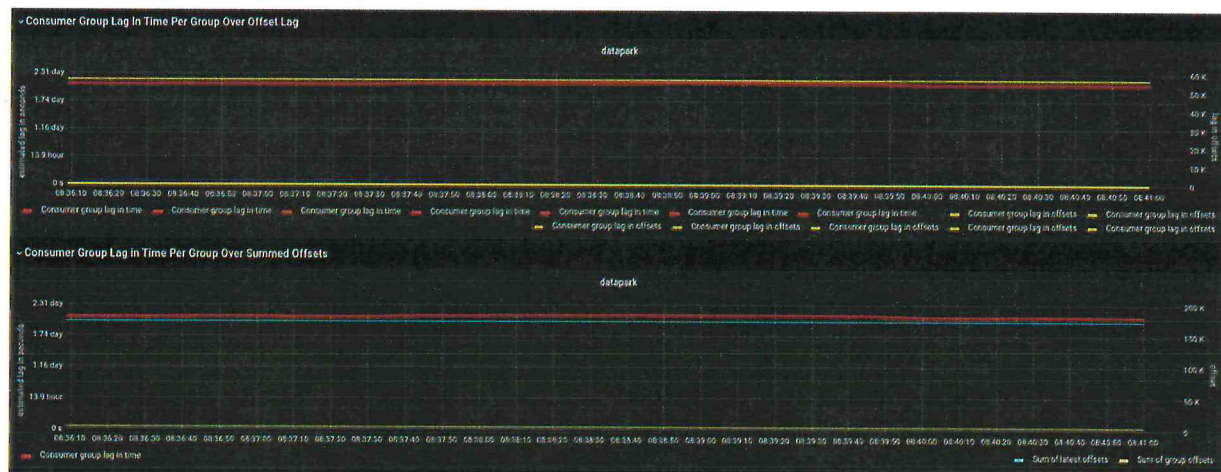


Рисунок 9. Задержка по времени для каждой группы потребителей

2.3 Мониторинг состояния кластера Greenplum

Информационная панель контроля нарушений в работе кластера Greenplum доступна в группе «General».

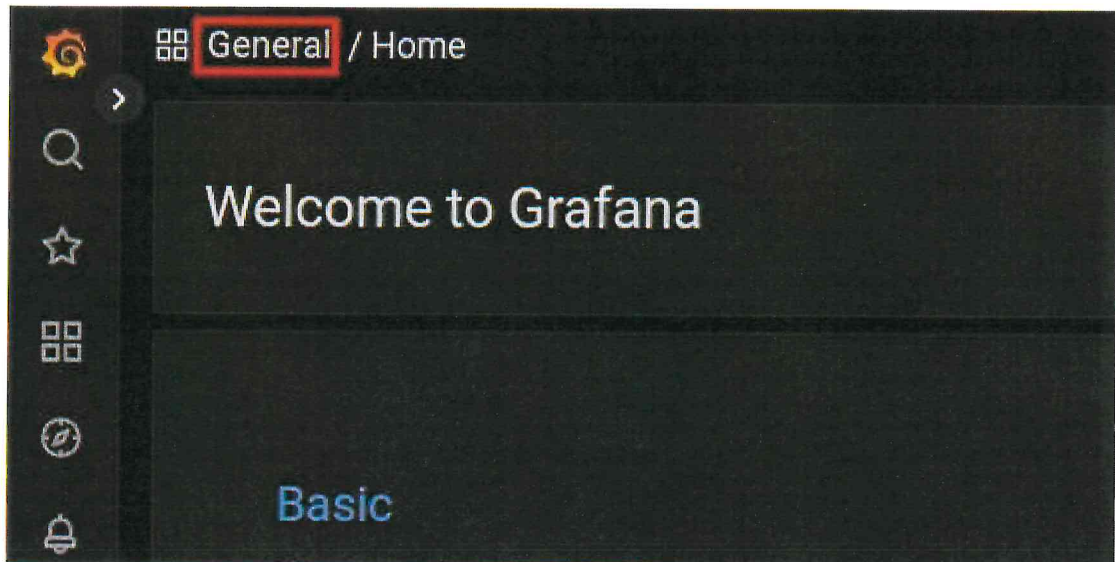


Рисунок 10. Расположение информационной панели

Для просмотра текущего состояния сегментов кластера и информации о подключениях, используется информационная панель «Greenplum Cluster».

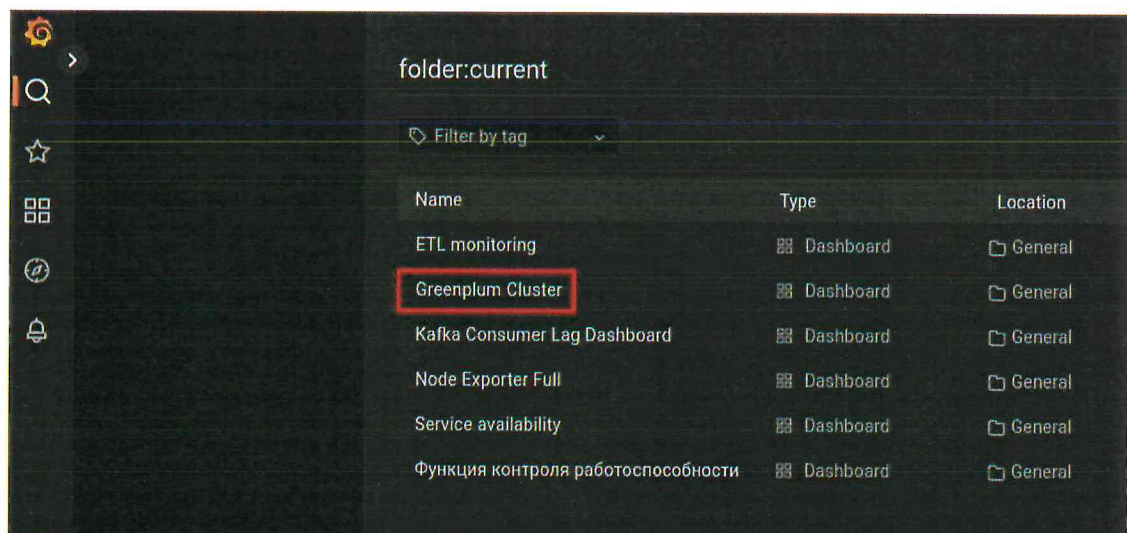


Рисунок 11. Мониторинг состояния кластера Greenplum

Информационная панель содержит следующие плитки:

1. «Количество зеркальных сегментов в роли основных» - метрика отображает количество сегментов, которые обычно являются зеркальными (резервными копиями основных сегментов), но в данный момент выполняют роль основных сегментов, возможно, из-за отказа исходных основных сегментов. Значение в нормальном состоянии - 0.

2. «Количество основных сегментов не имеющих зеркал» - показывает число основных сегментов данных, для которых не существуют зеркальные сегменты. Это может указывать на потенциальную точку отказа, так как отсутствие зеркала увеличивает риск потери данных при сбое соответствующего сегмента. Значение в нормальном состоянии - 0.

3. «Количество основных сегментов в роли зеркальных» - метрика указывает на количество сегментов, которые обычно функционируют как основные. Но временно, или по какой-либо причине, используются как зеркальные. Это может быть необычным состоянием и требовать дополнительного рассмотрения. Значение в нормальном состоянии - 0.

4. «Количество зеркальных сегментов в статусе DOWN» - сообщает о количестве зеркальных сегментов, которые в данный момент недоступны или не работают. Если таких сегментов много, это может сигнализировать о проблемах в системе и увеличивать риск потери данных. Значение в нормальном состоянии - 0.

5. «Активные основные сегменты» - общее количество основных сегментов, которые в настоящий момент функционируют и обрабатывают запросы. Это показатель текущей работоспособности кластера с точки зрения доступности данных. В нормальном состоянии - значение должно равняться количеству сегментов, предопределенному конфигурацией.

6. «Активные зеркальные сегменты» - число активных зеркальных сегментов, доступных для предоставления отказоустойчивости и обеспечения резервного копирования основных сегментов. В нормальном состоянии, значение должно равняться количеству активных основных сегментов.

7. «Uptime» - продолжительность непрерывной работы кластера Greenplum с момента последней перезагрузки или запуска. Чем больше время безотказной работы, тем более стабильной может считаться система.

8. «Количество подключений» - метрика отображает количество клиентских подключений к кластеру Greenplum. Может использоваться для мониторинга нагрузки на систему и определения пиковых периодов активности. Приводится информация о части подключений.

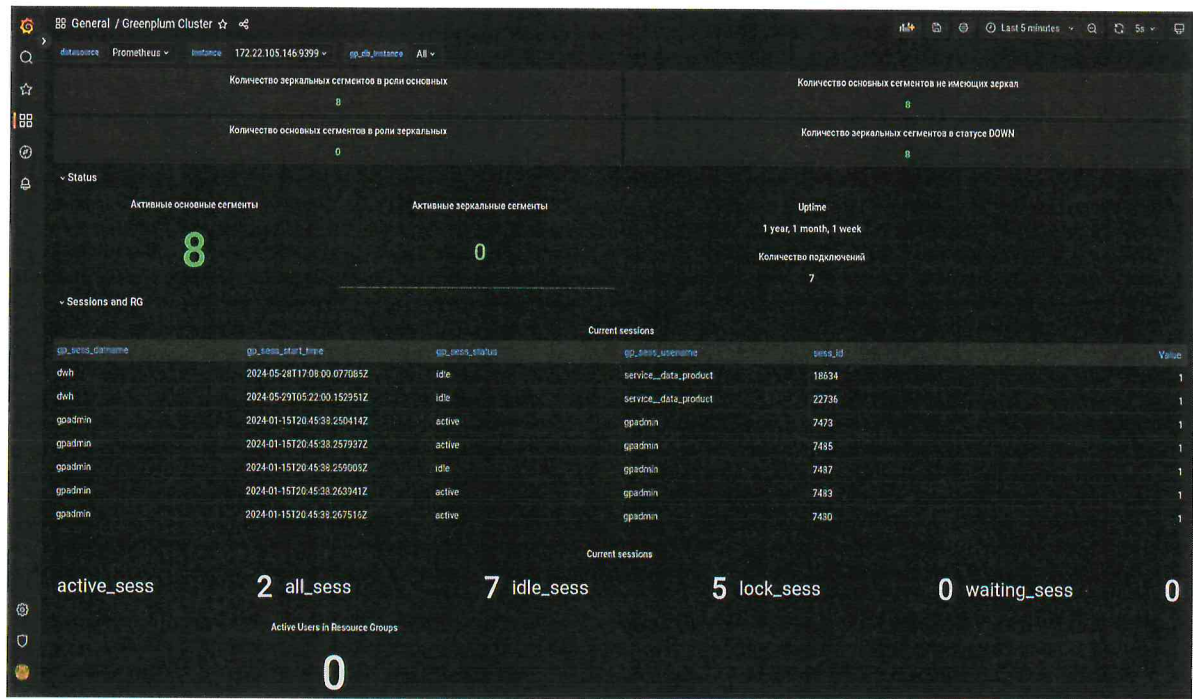


Рисунок 12. Контроль нарушений в работе кластера Greenplum

2.4 Функция контроля работоспособности

Для просмотра текущего состояния компонентов Apache Kafka, HDFS, YARN, PXF, Greenplum используется информационная панель «Функция контроля работоспособности», которая доступна в группе «General».

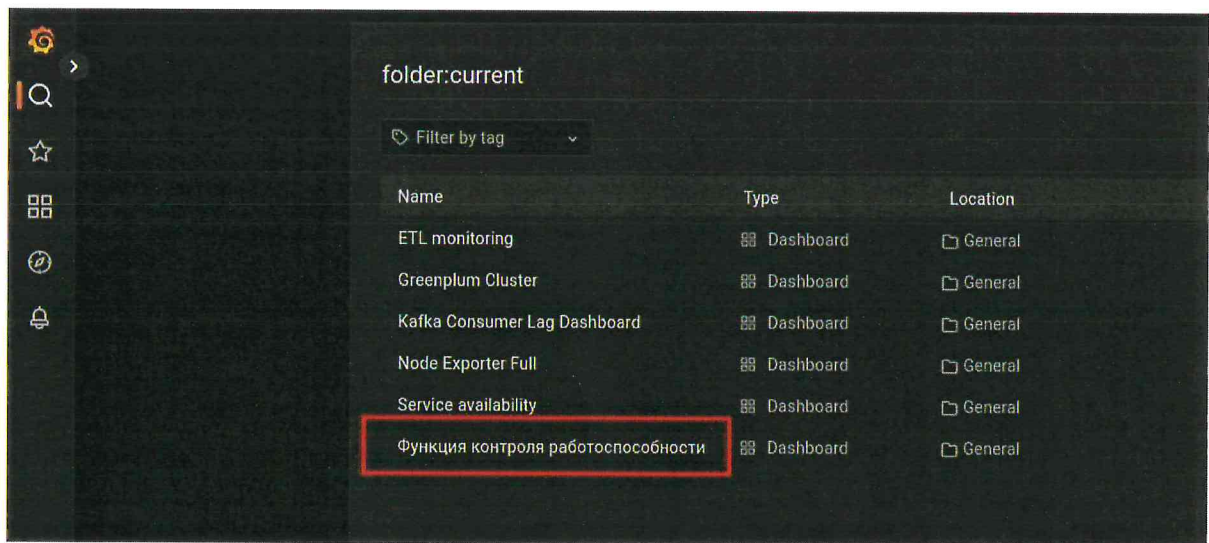


Рисунок 13. Информационная панель «Функция контроля работоспособности»

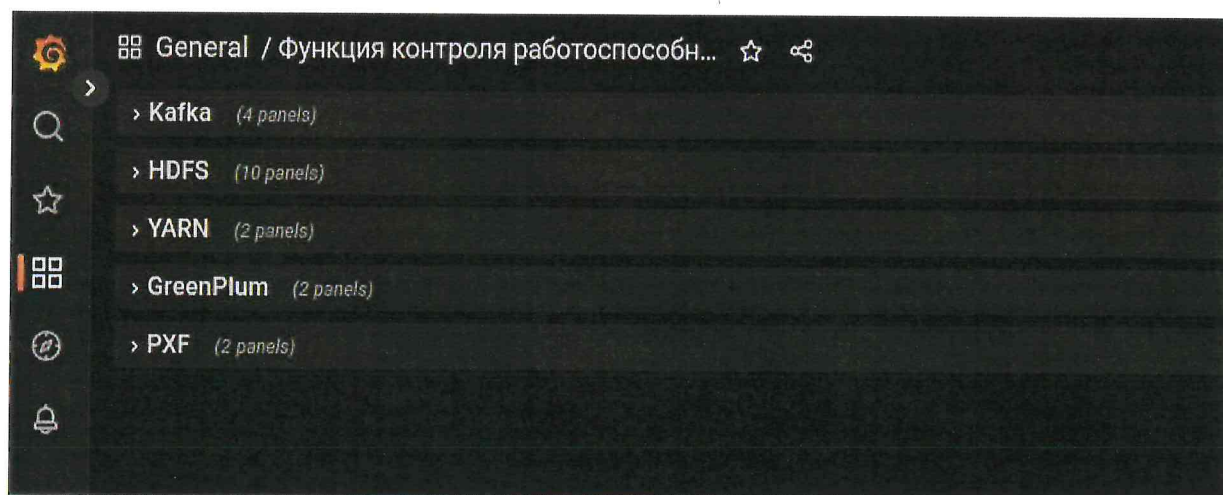


Рисунок 14. Компоненты

Для просмотра текущих показателей, необходимо нажать на стрелку слева от названия интересующего компонента.

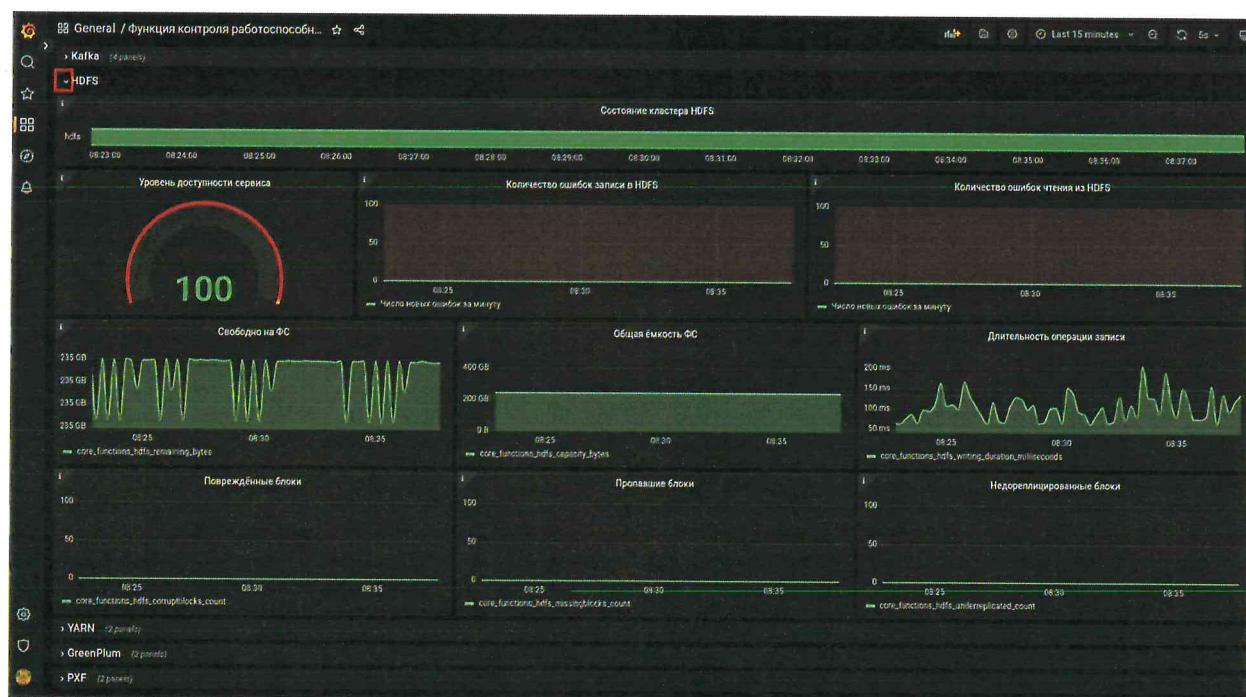


Рисунок 15. Состояние кластера HDFS

3 Действия пользователей при возникновении ошибок

При возникновении ошибок во время использования или необходимости консультации необходимо обратиться в техническую поддержку по электронной почте khd-support@myservicesdigital.ru или khd-support@myservices.digital.

Режим работы: обращения в техническую поддержку принимаются в будние дни с 9.00 до 18.00 часов.