|  |  |
| --- | --- |
|  |  |

**Руководство по техническому обслуживанию**

«Система управления знаниями AI-Solver»

Страниц: 13

Москва

2025**История изменений**

|  |  |  |  |
| --- | --- | --- | --- |
| Версия | Дата | Комментарии | Автор |
| 1.0 | 14.08.2025 | Первая версия документа | Сиглаев С.Н. |
|  |  |  |  |

# Аннотация

Настоящий документ содержит сведения, необходимые для эксплуатации информационной системы «Система управления знаниями «AI-Solver» в рамках применения средств программного обеспечения для реализации функций системного администрирования и технического обслуживания.

Документ подготовлен в соответствии с РД 50-34.698-90 - в части структуры и содержания документов, и в соответствии с ГОСТ 34.201-89 - в части наименования и обозначения документов.

**Содержание**

[Аннотация 2](#_Toc206087547)

[1 ВВЕДЕНИЕ 5](#_Toc206087548)

[2 ОБЩИЕ УКАЗАНИЯ 6](#_Toc206087549)

[2.1 ПОРЯДОК ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ 6](#_Toc206087550)

[2.2 ТРЕБОВАНИЯ К ПЕРСОНАЛУ 6](#_Toc206087551)

[3 ТРЕБОВАНИЯ К ТЕХНИЧЕСКИМ СРЕДСТВАМ 7](#_Toc206087552)

[4 ОПИСАНИЕ ФУНКЦИЙ 9](#_Toc206087553)

[4.1 ОБНОВЛЕНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ 9](#_Toc206087554)

[4.2 ВЫПОЛНЕНИЕ РЕЗЕРВНОГО КОПИРОВАНИЯ 10](#_Toc206087555)

[4.3 ПРОВЕРКА РАБОТОСПОСОБНОСТИ СИСТЕМЫ 10](#_Toc206087556)

[4.3.1 МОНИТОРИНГ РАБОТЫ СИСТЕМЫ 10](#_Toc206087557)

[4.3.2 ПРОВЕРКА РАБОТОСПОСОБНОСТИ ПРИ ПОМОЩИ LOG ФАЙЛОВ СИСТЕМЫ 11](#_Toc206087558)

[4.4 ВОССТАНОВЛЕНИЕ РАБОТОСПОСОБНОСТИ СИСТЕМЫ ПРИ АВАРИЙНЫХ СИТУАЦИЯХ 13](#_Toc206087559)

# 

# 1 ВВЕДЕНИЕ

Данный документ представляет собой руководство по техническому обслуживанию информационной системы «Система управления знаниями «AI-Solver» (далее также Система).

Документ предназначен специалисту по техническому обслуживанию, администратору баз данных и администратору Системы на этапе сопровождения.

При осуществлении регламентных работ также следует руководствоваться следующими документами:

* Руководство администратора (Технологическая инструкция) – далее также «Руководство»;
* Документация по Minio - <https://min.io/docs/minio/linux/index.html>
* Документация по Apache Solr - <https://solr.apache.org/guide/solr/9_8/index.html/>
* Документация по PostgreSQL - <https://www.postgresql.org/docs/14/index.html>
* Документация по Apache Nifi 1.28 <https://nifi.apache.org/documentation/v1/>
* Документация по Apache Nifi Registry <https://nifi.apache.org/docs/nifi-registry-docs/>
* Документация по Apache Zookeeper <https://zookeeper.apache.org/doc/r3.9.1/index.html>
* Документация по Apache Kafka 3.7.0 <https://kafka.apache.org/37/documentation/>
* Документация по Kafka UI v.0.7.2 <https://docs.kafka-ui.provectus.io/>
* Документация по RabbitMQ 4.0 <https://www.rabbitmq.com/docs/4.0>
* Документация по Redis <https://redis.io/docs/latest/>
* Документация по Qdrant <https://qdrant.tech/documentation/>
* Документация по Keycloack <https://www.keycloak.org/documentation>
* Документация по Docker Swarm <https://docs.docker.com/engine/swarm/>

Система предназначенна для сбора, хранения, анализа и распространения структурированных и неструктурированных знаний внутри организации. Цель системы — повышение эффективности работы и поиска решений за счёт централизованного доступа к корпоративным знаниям, автоматизации обработки данных и интеллектуального анализа информации.

# 2 ОБЩИЕ УКАЗАНИЯ

## 2.1 ПОРЯДОК ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ

В рамках сопровождения Системы реализуется выполнение следующих функций:

* обновление программного обеспечения;
* выполнение резервного копирования;
* осуществление проверки штатными средствами работоспособности системы;
* восстановление работоспособности системы при аварийных ситуациях.

Требования к уровню подготовки специалиста по техническому обслуживанию

Системы приведены в разделе 2.1 Руководства и в п.2.2 настоящего документа.

## 2.2 ТРЕБОВАНИЯ К ПЕРСОНАЛУ

Выполнение функций, указанных в разделе 2.1, предполагает, соответствие профессионального уровня специалиста по техническому обслуживанию и администрированию Системы следующим требованиям:

* высокий уровень квалификации и практического опыта выполнения работ по сопровождению баз данных;
* умение установки, модернизации, настройки параметров программного обеспечения (СУБД);
* наличие опыта разработки, управления и реализации эффективной политики доступа к информации, хранящейся в прикладных базах данных;
* высокий уровень квалификации и практического опыта выполнения работ по модернизации, настройке и мониторингу работоспособности комплекса технических средств (серверов, рабочих станций);
* профессиональные знаниями в области конфигурирования и настройка программно-технических средств Системы;
* навыки по диагностике типовых неисправностей, замене базовых узлов периферийных устройств, имеющих ограниченный ресурс, настройке локальной компьютерной сети, контроля доступа к сетевым ресурсам.

# 3 ТРЕБОВАНИЯ К ТЕХНИЧЕСКИМ СРЕДСТВАМ

Требования к показателям назначения Системы: потребителями информации

Системы могут быть до 500 одновременных пользователей.

Предполагаемый уровень нагрузки на систему оценивается как:

* количество загружаемых документов в систему за месяц: до 15 млн. шт.;
* количество загружаемых файлов в систему за месяц: до 1 млн шт. (в зависимости от типа и размера файла)
* пиковая интенсивность поисковых запросов: до 100 запросов в секунду
* пиоквая интенсивность загрузки документов из системы: до 60 документов в минуту

Для обеспечения целевых показателей, минимальный набор комплекса технических средств, обеспечивающих работу Системы, должен состоять из следующих основных элементов:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Устанавливаемы сервисы | ОС | RAM | Disk | Примечание |
| 1 | registry  db  solr-node-1  zookeper | Ubuntu 22.04 | 16GB | 1TB | Стабильная сеть с пропускной способностью не менее 1000 Мбит/с |
| 2 | keycloack  redis  search-backend  search-frontend  service-call-analyzer  view-server | Ubuntu 22.04 | 16GB | 200GB | Стабильная сеть с пропускной способностью не менее 1000 Мбит/с |
| 3 | nifi  nifi-registry  zookeeper  kafka  kafka-ui  file-processor  rabbitmq  ocr-server  ocr-worker | Ubuntu 22.04 | 16GB | 200GB | Стабильная сеть с пропускной способностью не менее 1000 Мбит/с |
| 4 | stt-server  stt-worker  stt-worker-punctuator  ner-server  ner-worker-1  ner-worker-2  ner-worker-3 | Ubuntu 22.04 | 32GB | 200GB | Стабильная сеть с пропускной способностью не менее 1000 Мбит/с |
| 5 | insight-ai-server  insight-ai-indexer  qdrant  redis  rabbitmq  text-gen-web-ui | Ubuntu 22.04 | 16GB | 200GB | Требуется видеокарта для запуска LLM (не менее 24GB VRAM) |

Для увеличения пиковой интенсивности запросов необходимо:

* Увеличить размеры heap памяти Apache Solr
* Использовать шардирование для больших коллекций
* Настройка репликации для балансировки нагрузки
* Использование SSD

Для увеличения количества обрабатываем файлов необходимо увеличение количества воркеров (OCR, STT, NER)

# 4 ОПИСАНИЕ ФУНКЦИЙ

## 4.1 ОБНОВЛЕНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

* Установка программ и настройка операционной системы и общесистемных компонентов осуществляются стандартным образом.
* Создайте резервную копию всех данных (см. раздел 4.2).
* Проверьте доступные обновления для каждого компонента (например, используйте docker pull для новых образов).
* Внимательно изучите журнал изменений (changelog) для выявления возможных несовместимостей.
* После обновления выполните проверку системы (см. раздел 4.3).

## 4.2 ВЫПОЛНЕНИЕ РЕЗЕРВНОГО КОПИРОВАНИЯ

В системе все тома примонтированы к директории /share/volumes, что упрощает процесс резервного копирования.

Резервное копирование томов необходимо для:

* Восстановления данных после сбоя – при повреждении или потере данных.
* Миграции приложений на другой сервер – перенос данных между серверами Swarm.
* Аварийного восстановления (Disaster Recovery) – восстановление работы системы после серьезных инцидентов.

Резервные копии должны храниться вне основного сервера (например, S3, NFS, локальный архивный диск).

## 4.3 ПРОВЕРКА РАБОТОСПОСОБНОСТИ СИСТЕМЫ

### 4.3.1 МОНИТОРИНГ РАБОТЫ СИСТЕМЫ

Мониторинг работы системы осуществляется с помощью контроля связанных с компонентами системы метрик. В основе контроля лежат:

* Сбор данных о состоянии сервиса в реальном времени;
* Анализ трендов производительности;
* Настройка предупреждений (оповещений) при превышении или снижении определённых пороговых значений.

Ключевые метрики для контроля

a) Производительность системы:

* Утилизация процессора (CPU Usage) — процент времени, когда процессор системы активен;
* Использование оперативной памяти (Memory Usage) — объём занятой и освобождённой памяти;
* Использование места на диске (Disk Usage) — свободный и занятый объём памяти системы.

b) Работа сети:

* Пропускная способность сети (Network Bandwidth) — объём переданных данных;
* Количество сетевых запросов и пакетов (например, TCP/IP).

c) Загрузка приложений:

* Количество запросов в секунду (Requests per Second, RPS);
* Среднее время отклика (Response Time);
* Уровень ошибок (Error Rate) — процент неудавшихся запросов.

d) Специфичные метрики по компонентам:

Состояние базы данных (например, PostgreSQL):

* Количество активных подключений;
* Время выполнения запросов.

Сервисы хранения (MinIO):

* Успешность выполнения операций ввода-вывода (IOPS);
* Производительность чтения/записи.

Индексация данных (Apache Solr):

* Время индексации новых документов;
* Скорость поиска (Search Latency).

### 4.3.2 ПРОВЕРКА РАБОТОСПОСОБНОСТИ ПРИ ПОМОЩИ LOG ФАЙЛОВ СИСТЕМЫ

Логирование работы Системы осуществляется штатными средства серверов. Информация о полученных запросах и ответных сообщениях сохраняется в лог файл. Все события фиксируются в лог файлах серверов. Осуществить просмотр лог файлов можно стандартными средствами Linux.

Логи — это текстовые или структурированные файлы, содержащие записи об активности системы. Они помогают выяснить причины неисправностей или подозрительного поведения, а также проводить аудит системы.

Типы логов

a) Системные логи:

* Логи операционной системы (syslog): сообщения ядра, драйверов и ОС;
* Логи сервисов: записи активности сервисов и их взаимодействия.

b) Логи приложений:

* Логи запросов и ответов (например, веб-серверов, REST API);
* Логи ошибок: причину сбоев, ошибки ввода-вывода, проблемы с подключениями к базе данных и непредвиденные исключения.

c) Логи безопасности:

* Записи аутентификации пользователей (успешные/неудачные попытки входа);
* Логи авторизации (Keycloak) — информация о действиях пользователей в системе;
* Логи о подозрительных действиях, например, попытках взлома.

# 4.4 ВОССТАНОВЛЕНИЕ РАБОТОСПОСОБНОСТИ СИСТЕМЫ ПРИ АВАРИЙНЫХ СИТУАЦИЯХ

Нарушение условий выполнения технологического процесса проявляется в виде невозможности выполнения и/или завершения технологической операции.

Причиной нарушения условий выполнения технологического процесса, как правило, являются сбои в аппаратном и программном обеспечении.

Если авария вызвана сбоем работы оборудования, нужно определить причину сбоя, устранить сбойный элемент оборудования и протестировать систему.

В случае если произошел критический сбой, например выход из строя системного жесткого диска, необходима процедура восстановления из резервной копии.

Если авария вызвана сбоем работы программного обеспечения, нужно определить в каком именно программном модуле возникла ошибка и исправить ее.

Для определения источника ошибок необходимо использовать журналы событий (log-файлы).

В случае отказа работы сервера приложений, необходимо осуществить восстановление работы сервера. При необходимости провести восстановление метаданных из резервной копии.

В случае отказа системы хранения данных, необходимо осуществить восстановление работы системы, при необходимости провести восстановление из резервной копии.