

INFATON Chat

Установка и администрирование

Версия 17.11.419-registry.39 · Редакция 15 октября 2026 года

Инструкция предназначена администратору серверной установки INFATON Chat версии 17.11.419-registry.39. Она описывает подготовку Linux-сервера, первый запуск, резервирование, восстановление и обслуживание. Перед выполнением проверьте, что указанные сервисы и средства входят в паспорт и манифест вашего комплекта. Пакет обновления только ядра не заменяет полный установочный комплект.

Подготовка среды

Нужны Linux amd64, Docker Engine, Compose v2, Python 3.11 или новее, shell, openssl, curl и средства распаковки архива. У администратора должны быть права на Docker. Проверьте, что его команды, установщик и служба обновления обращаются к одному Docker daemon; особенно важно проверить это при использовании sudo и rootless Docker. Для входа пользователей настройте HTTPS-прокси, DNS и сертификат. СУБД и внутренние сервисы не публикуйте в пользовательскую сеть.

Серверная поставка предназначена для Linux x86_64 (amd64), Docker Engine и Docker Compose v2. Совместимость с конкретным дистрибутивом Linux и его версией проверяют перед установкой. Ресурсы зависят от числа пользователей, документов, выбранных профилей и модели. Для предварительного планирования можно использовать ориентиры: базовый состав без базы знаний — от 2 ядер, 4 ГБ памяти и 60 ГБ диска; локальные эмбединги — от 16 ГБ памяти и дополнительного диска. Это предварительные ориентиры, а не подтверждённый минимум и не норматив производительности. Перед установкой определите полный состав профилей, сервисов и моделей, рассчитайте ресурсы именно для него и предусмотрите место для данных, точных образов, распаковки и резервных копий. Основная языковая модель рассчитывается отдельно.

Получение комплекта

Получите архив, паспорт версии, контрольную сумму и инструкцию проверки подписи именно для этого вида поставки. Сверьте SHA-256 и фактический состав с манифестом до распаковки. Контрольная сумма, взятая из того же изменяемого источника, сама по себе не подтверждает происхождение; при несовпадении суммы или подписи не используйте комплект. Подпись пакета обновления ядра не распространяется на полный архив установки: для каждого вида поставки применяйте его собственный паспорт и способ проверки.

Сверяйте фактические сервисы, образы, модели и сопроводительные материалы с паспортом и манифестом именно полученного комплекта. При наличии моделей или исходных материалов их точный состав и контрольные суммы задаются манифестом; не дополняйте комплект файлами из другой версии. Уведомления о сторонних компонентах должны находиться в предусмотренном комплектом разделе.

Настройка перед первым запуском

Распакуйте комплект в отдельную папку. Скопируйте .env.template в .env, если файл ещё не существует. Не подставляйте чужие секреты: пустые POSTGRES_PASSWORD и ERPTWIN_SECRETS_KEY заполняет установщик. Сохраните .env с правами 0600. Для штатного пользовательского режима используйте ACCESS_MODE=strict и включённую проверку CSRF. Режим ACCESS_MODE=off не предназначен для доступа рабочих пользователей.

Параметр	Назначение
BOX_PROJECT	Уникальное имя установки в Docker Compose; по умолчанию erptwin-box
HTTP_PORT	Свободный локальный порт, по умолчанию 8080
BIND_ADDR	Оставьте 127.0.0.1 при доступе через HTTPS-прокси
APP_URL	Фактический HTTPS-адрес этой установки для внешних ссылок
COMPOSE_PROFILES	Профили из комплекта: analytics,rag,voice; сохраняйте выбранный состав

Установка и первый вход

До запуска проверьте, какой Docker daemon выбран и какие установки уже работают на сервере. Установщик загружает образы, создаёт недостающие секреты и запускает проект. Если Docker отсутствует, подготовительный скрипт может устанавливать пакеты ОС. В закрытом контуре подготовьте зависимости заранее. На общем сервере не допускайте изменения служб соседних приложений.

```
docker info --format "{{.ID}}"
bash ./install.sh --no-system-updater
docker compose -f docker-compose.box.yml ps
```

Команды выполняются под пользователем с правами на выбранный daemon и каталог установки. Для обычного системного Docker на выделенном сервере могут потребоваться административные права. При rootless Docker используйте его CLI, DOCKER_HOST и владельца процесса; обычный sudo может выбрать другой daemon.

Флаг --no-system-updater запускает выбранный комплект без системного автообновлятора. Этот режим использован при проверке установки версии .39. Если требуется включить системное обновление, применяйте инструкцию именно из принятого комплекта и заранее проверьте, к какому Docker daemon она обращается. Не применяйте административные команды, если они переключают установку на другой daemon.

До запуска проверьте штатным средством комплекта целостность всех вложенных файлов модели, включая ONNX, и ожидаемые repository/revision. Не дополняйте неполный набор файлами случайной версии. При наличии вложенных весов bge-m3 установщик добавляет профиль rag к выбранным профилям. Первый ответ ядра ещё не означает завершение загрузки всех моделей. Убедитесь, что выбранные сервисы запущены и не завершаются с ошибкой.

Откройте /setup по адресу своей установки. Сначала получите идентификатор установки и загрузите лицензию, выданную для этого идентификатора. Затем зарегистрируйте владельца с собственным паролем. Лицензия от другого экземпляра может быть отклонена, даже если её срок ещё не истёк. HTTP 402 до активации отличается от ошибки установки.

Подключите основную модель: адрес OpenAI-совместимого API, идентификатор модели и необходимые учётные данные. Запрос и передаваемый модели контекст направляются на этот адрес; при удалённом endpoint данные могут покидать инфраструктуру организации. Для автономной работы используйте локально размещённую модель и заранее доступные файлы и зависимости, а затем проверьте сетевые маршруты и интеграции. Для 1C укажите адрес опубликованного MCP-моста и учётную запись с согласованными правами. Подключение 1C можно настроить в мастере или в разделе «Подключение 1C». Настройки модели доступны через «Настройки → LLM».

Проверка после настройки

1. Откройте /healthz и сверьте версию; /readyz должен подтверждать доступность БД.

2. Войдите в браузер через HTTPS. Выполните простой запрос к модели и проверьте фактический ответ.
3. Выполните разрешённое чтение из нужной базы 1C. Сравните результат с источником.
4. Для каждого используемого сервиса проверьте отдельный пример: файл для OCR, документ базы знаний, голосовую запись, вычисление.
5. Сохраните результат, выйдите и откройте его повторно. Проверьте права обычного пользователя отдельно от владельца.

Если приложение не запускается, прочитайте журнал своего проекта. Не выполняйте глобальный перезапуск Docker или WSL, когда на сервере работают другие установки.

```
docker compose -f docker-compose.box.yml logs --tail=100 app
```

Перед передачей журнала поддержке удалите пароли, токены и содержание документов. Не отключайте проверку TLS для устранения ошибки сертификата: настройте доверенную цепочку и правильное имя сервера.

Резервное копирование

Из каталога установки под пользователем выбранного Docker daemon выполните резервирование на защищённый носитель. На время согласованного снимка инструмент останавливает службы только выбранного проекта и затем возвращает ранее работавшие службы. Исключите стороннюю запись непосредственно в PostgreSQL на время снимка.

```
python3 backup-box.py --to /secure-backup
python3 backup-box.py --list --to /secure-backup
python3 backup-box.py --verify /secure-backup/infaton-backup-20261015T120000Z-0123abcd.tar.gz
```

В примере проверки замените имя архива на точное имя, показанное вашей установкой.

Для проверки укажите точный путь к архиву, который напечатал инструмент или показывает список копий. Полная копия охватывает базу данных, конфигурацию, файлы установки и файловые тома, сохранённые права доступа (включая ACL), а также ключи и секреты, необходимые для чтения данных. Состав и контрольные суммы фиксируются манифестом. Архив содержит секреты, не зашифрован и не подписан; ограничьте доступ к нему и храните его на защищённом носителе. Не прикладывайте архив к обычному обращению в поддержку.

Образы, исходные файлы моделей, внешняя 1C, основная модель, ОС и HTTPS-прокси не заменяются этой копией. Храните точный исходный комплект отдельно и резервируйте внешние системы по их регламентам. Для нестандартных томов сначала проверьте полноту резервирования. Старый SQL-дамп не является полной копией установки.

Восстановление в новый проект

Распакуйте точный исходный пакет .39 в новую папку. Убедитесь, что в ней нет действующего Compose-проекта и что точные образы из манифеста доступны локально. Запустите CLI из каталога этой новой установки, задайте новое имя Compose и свободный порт. Штатное восстановление не предназначено для перезаписи действующего проекта.

```
python3 backup-box.py --restore /secure-backup/infaton-backup-20261015T120000Z-0123abcd.tar.gz \
--project infaton-restored --port 8081
```

В примере восстановления замените имя архива на точное имя проверенной копии.

До переключения пользователей проверьте вход, привязку лицензии, файлы, историю, права и расшифровку подключений. Проверьте реальный обмен с моделью и 1C. Контрольные суммы подтверждают целостность файлов, но не заменяют пробное восстановление и проверку сценария.

При неуспехе сохраните журнал и .restore-incomplete.json. Не удаляйте маркер ради продолжения поверх незавершённого восстановления. Устраните причину и действуйте по РЕЗЕРВНАЯ-КОПИЯ.md, включённой в комплект.

Обновление и завершение эксплуатации

Перед обновлением зафиксируйте текущую версию и состав, создайте и проверьте полную резервную копию. Получите подписанный пакет обновления именно для этой поставки, сверьте его манифест и контрольные суммы и следуйте инструкции, вложенной в комплект. Штатное обновление сохраняет базу данных, файлы и ACL, конфигурацию и связанные ключи; архив обновления не заменяет резервную копию. После применения сверьте версию, вход, данные, права, файлы и нужный рабочий сценарий. При неуспехе сохраните журналы и действуйте по согласованному плану восстановления.

Для временной остановки из каталога установки выполните команду ниже. Она относится к проекту, указанному в .env. Возобновление производится командой `up -d`.

```
docker compose -f docker-compose.box.yml stop
docker compose -f docker-compose.box.yml up -d
```

Не удаляйте тома при обычной остановке. Окончательное удаление данных выполняется по решению владельца после проверки архивов и сроков хранения.

Управление нагрузкой

LLM_MAX_CONCURRENT ограничивает одновременные запросы к одному сочетанию адреса, модели и ключа внутри одного процесса приложения. По умолчанию 6; допустимы целые значения от 1 до 256. Это не распределённая квота для нескольких процессов. Подбирайте значение по возможностям модельного сервера.

Плановые фоновые оценки уступают интерактивным запросам. SELF_EVAL_MODE задаёт режим: full - обычный, light - облегчённый, off - без планового цикла. Эти параметры не гарантируют качество ответа или время выполнения задачи. Изменения .env требуют применения к соответствующему сервису; согласуйте окно обслуживания.

Помощь администратора

Поддержка: info@infaton.ru, +7 991 006-61-00. Передайте версию, время, конфигурацию сервисов без секретов, ожидаемый результат и воспроизводимые действия. Укажите, затронуты ли данные и доступно ли восстановление. Объём поддержки, сроки и условия обновлений определяются договором поставки или сопровождения.