

Платформа Эра

Российская информационно-коммуникационная платформа

Отчет по нагрузочному тестированию

Оглавление

ОГЛАВЛЕНИЕ	1
1. ВВЕДЕНИЕ	3
1.1. Назначение документа	3
1.2. Цели тестирования	3
1.3. Объекты тестирования	3
2. КОЛИЧЕСТВЕННЫЕ ПОКАЗАТЕЛИ	3
3. ВЫВОДЫ	3
4. ПОДРОБНОСТИ ТЕСТИРОВАНИЯ И РЕКОМЕНДАЦИИ	4
5. САЙЗИНГ ЦЕЛЕВОЙ СИСТЕМЫ	4
6. МЕТОДИКА ПРОВЕРКИ	6
6.1. Соответствие методике проверки	6
6.1.1. Эмулятор PSTN	6
6.1.2. Эмулятор операторов	8
6.1.3. Целевая система	9
6.2. Сбор и анализ данных	10
6.2.1. Требования по утилизации аппаратных ресурсов	10
6.2.2. Отчетность контакт-центра	10
7. ОБЗОР СТЕНДА	11
7.1. Характеристики стенда	11
7.1.1. Целевая система	11
7.1.2. Эмуляторы	11
7.2. Приложение мониторинга	12
7.3. Подключение тестовых телефонов	13
8. ПРОВЕДЕНИЕ НАГРУЗОЧНОГО ТЕСТА	14
9. МЕТРИКИ И АНАЛИЗ ТЕХНИЧЕСКИХ РЕЗУЛЬТАТОВ	14
9.1. Утилизация ресурсов	14
9.1.1. Утилизация CPU (процессоров)	14
9.1.2. Утилизация оперативной памяти	16
9.1.3. Утилизация жестких дисков	17
9.1.4. Объем обмена данными с жесткими дисками	18
9.1.5. Количество операций к жестким дискам	20
9.1.6. Объем занятого места на диске	21
9.1.7. Сетевой трафик	22
9.1.8. AVG (показатель загруженности операционной системы Linux)	23
9.1.9. Утилизация CPU микросервисами коммуникационной платформы	25
9.1.10. Утилизация оперативной памяти микросервисами коммуникационной платформы	26

9.2.	Записи разговоров	27
9.3.	Остановка теста	28
9.4.	Счетчики	35
9.5.	Метрики	35

1. Введение

1.1. Назначение документа

Предоставить детальный анализ результатов нагрузочного тестирования Платформы ЭРА, в соответствии с предоставленной методикой.

1.2. Цели тестирования

- Проверка технической готовности программного обеспечения Платформа ЭРА к обработке существенного объема вызовов;
- Способность системы работать без деградации сервиса.

1.3. Объекты тестирования

- Работа модулей мониторинга и управления, компоненты системы, отвечающие за прием входящих звонков и совершение массового исходящего обзвона.
- Система маршрутизации и распределения вызовов.
- Система записи.

2. Количественные показатели

- 620 одновременно работающих операторов. Дополнительно 10 операторов для возможности ручной проверки качества звука и корректности алгоритмов маршрутизации;
- 72 часа непрерывной нагрузки;
- принято 1,5 миллиона входящих вызовов;
- совершено 18,5 миллионов исходящих вызовов, CPS 80+ вызовов в секунду;
- обработано операторами 2,7 миллиона вызовов.

3. Выводы

Тест пройден успешно. Ограничений для масштабирования не выявлено. У Платформы Эра имеется существенный запас производительности. Система готова к увеличению нагрузки на имеющихся серверах.

Вывод о наличии значительного запаса производительности сделан на основании анализа загруженности системных ресурсов. В частности - ядер центрального процессора (от 15 до 45%, в среднем - 35%) и оперативной памяти (в пределах 25%).

Эти показатели позволяют утверждать, что увеличение нагрузки в два раза с незначительным перераспределением микросервисов между серверами позволят остаться в рамках описанных Методикой показателей (70% CPU и 80%RAM).

Ключевые факторы, влияющие на загрузку системных ресурсов — это количество одновременных разговоров, количество сессий IVR и частота вызовов CPS. Микросервисная архитектура предполагает горизонтальное масштабирование. Препятствий для кратного увеличения нагрузки не выявлено.

Таким образом, можно говорить о потенциальной возможности увеличения CPS входящих вызовов с 6 до 12 вызовов в секунду, исходящих с 80 до 160 вызовов в секунду, операторов с 620 до 1240 без наращивания вычислительных мощностей текущего нагрузочного стенда.

4. Подробности тестирования и рекомендации

1. Платформа была развернута на 4х виртуальных серверах с относительно небольшим сайзингом, расположенных в Яндекс.Облаке. Дополнительно было использовано 2 сервера для эмуляции работы операторов и клиентов. Подробные характеристики стенда описаны в п.7.1.
2. Было совершено 18 565 353 исходящих вызовов за 72 часа.
3. Количество обработанных входящих вызовов – 1 555 083 шт.
4. Качество звука при вызовах нареканий не вызывало.
5. Утечек памяти на серверах со временем не выявлено.
6. Роста утилизации ресурсов со временем не выявлено.
7. Проблем со стабильностью работы системы за истекший период не выявлено.
8. Ни один микросервис за время теста не перезапускался.
9. Качество записей разговоров нареканий не вызывало. Входящие вызовы при распределении на операторов записывались. В случайный момент времени можно было обратиться к журналу звонков, обнаружить записанный разговор, прослушать запись. Выборочное прослушивание записей подтвердило высокий уровень качества разговоров и записей.
10. 1% потерянных вызовов связан с нехваткой операторов на второй линии в момент, когда размер очереди первой линии еще не слишком велик, чтобы операторы второй линии быстро освобождались. Это проявлялось в основном в моменты старта кампаний.
11. Облачная инфраструктура способна заметно ухудшить отклики и качество доставки пакетов, но система справляется успешно за счет распределения и параллелизации критических процессов.
12. Заметная загруженность жестких дисков СУБД при одновременно низком расходе оперативной памяти:
 - требует подготовки серверов под СУБД к увеличению нагрузки;
 - может быть преодолена:
 - распределением данных по разным экземплярам СУБД на разных серверах;
 - в части архивных данных переходом на колоночную БД для хранения архивных записей;
 - в части таблиц контрагентов распараллеливанием и мультиплексированием операций при загрузке колл-листов.

5. Сайзинг целевой системы

Микросервисная архитектура с горизонтальным масштабированием микросервисов в режиме active-active и равномерная балансировка нагрузки между однотипными сервисами дает осно-

вания утверждать, что разделение ресурсоемких микросервисов позволяет эффективно распределить нагрузку между несколькими серверами. При этом существует несколько механизмов балансировки нагрузки, наиболее часто используемый - hashring.

В ходе текущего тестирования использовалось 6 активных медиашлюзов, которые равномерно обрабатывали по 16.7% звонков. Соответственно, при увеличении нагрузки достаточно лишь увеличить их количество.

Среди других высоконагруженных сервисов - веб-сервер (ws), юзер-агент (b2b), движок сценариев (ivr) и модуль обработки данных (dms). Первые три масштабируются в любых количествах по механизму hashring, а четвертый (dms) имеет три уровня масштабирования - сначала по доменам, затем по классам внутри домена, и наконец по идентификаторам экземпляров класса.

При реализации балансировщиков на всех уровнях используются быстрые кешы, поэтому выбор конкретного экземпляра микросервиса занимает от 7 до 12 микросекунд, что ничтожно мало даже при тысячах операций в секунду, особенно с учетом многопоточности.

Текущее нагрузочное тестирование подтвердило, что архитектура и потенциальная производительность системы позволяют масштабировать нагрузку от 100 до 10.000+ операторов. Все перечисленные факторы позволяют ожидать линейного увеличения производительности системы при выделении дополнительных вычислительных мощностей.

На основании внутренних тестов и проведенного нагрузочного тестирования был рассчитан целевой сайзинг для 5600 одновременно работающих операторов (6600 одновременно работающих пользователей системы), 4100 портов одновременно работающих автоинформаторов, плюс 15000 портов с 800 наборами в секунду (CPS) для исходящих активностей:

Конфигурация	Необходимые ресурсы
Количество узлов	30
Физических ядер CPU, сумма	240
Гарантированная доля	100%
Объём памяти, сумма	480Gb
Дисковое пространство, сумма	3600Gb

Примечание:

В расчете не учтена отказоустойчивость в рамках распределения по различным ЦОД. Если необходима географически распределенная система, которая поддерживает работоспособность при полном отказе любого из ЦОД, то количество ресурсов необходимо увеличить кратно количеству ЦОД. При этом архитектура платформы Эра и подход к резервированию не требуют наличия второго плеча в режиме stand-by, а позволяют при необходимости задействовать часть его ресурсов в режиме active-active.

Примечание:

Подразумевается использование внешней СУБД и внешнего долговременного хранилища записей разговоров и экранов операторов.

6. Методика проверки

6.1. Соответствие методике проверки

С систем эмуляторов подается нагрузка с непрерывной плотностью **6 входящих вызовов в секунду**, обрабатываемая целевым стендом по алгоритмам обработки входящих вызовов с учетом вероятностных характеристик, приведенных в методике верификации для этого типа вызовов;

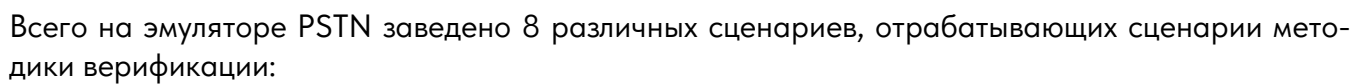
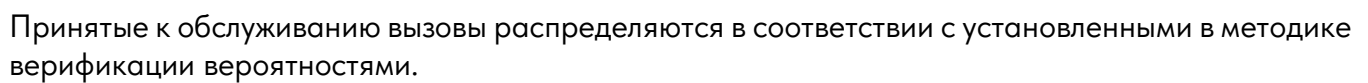
На целевой системе запущена исходящая кампания с плотностью **80 вызовов в секунду**.

Метрики приложения оператора проверяются по результатам работы эмулятора оператора.

Метрика	Описание	Значение
Подключение оператора	Время выполнения подключения оператора	90% не более 3 сек 10% не более 6 сек
Отключение оператора	Время выполнения отключения оператора	90% не более 3 сек 10% не более 6 сек
Ответ на звонок	Интервал времени между отправкой команды ответить на входящий звонок и получением события соединения	90% не более 1 сек 10% не более 2 сек
Отбой звонка	Интервал времени между отправкой команды отбоя звонка и получением события разъединения	90% не более 1 сек 10% не более 2 сек
Слепой перевод	Интервал времени между отправкой команды слепого трансфера звонка и получением события завершения перевода, в приложении оператора	90% не более 2 сек 10% не более 3 сек
Исходящий звонок	Интервал времени между отправкой команды исходящего звонка и получением события доставки звонка абоненту	90% не более 2 сек 10% не более 3 сек
Инициирование двухшагового перевода	Интервал времени между отправкой команды инициирования консультационного звонка и получением события доставки звонка третьему абоненту	90% не более 2 сек 10% не более 3 сек
Завершение двухшагового перевода	Интервал времени между отправкой команды завершения перевода и получением события завершения консультационного звонка	90% не более 1 сек 10% не более 2 сек

6.1.1. Эмулятор PSTN

Обслуживание исходящих вызовов производится в системе-эмуляторе PSTN с помощью сценариев IVR с заданным в методике верификации распределением. 92% - отбой, 8% - на обслуживание.

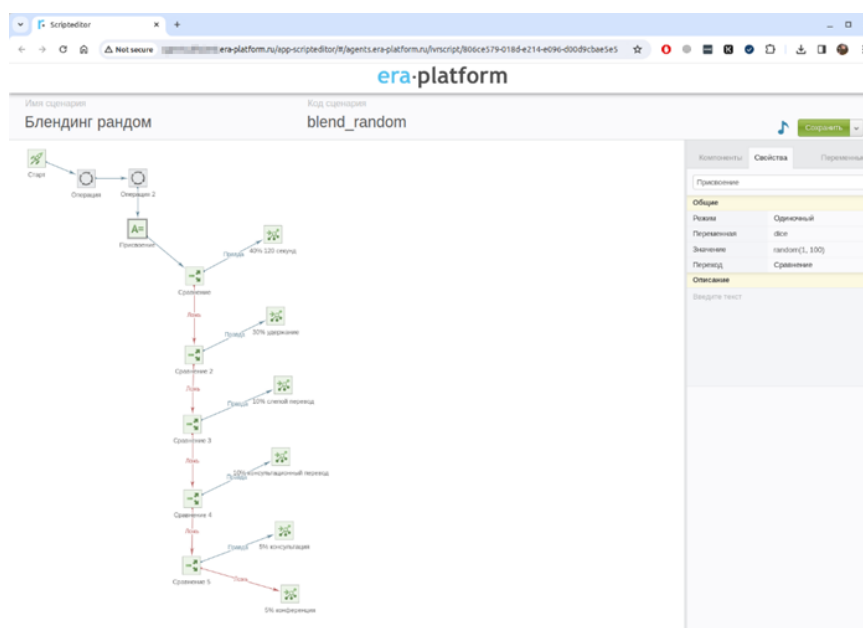


The screenshot shows the 'Era: Настройки' application interface. The top navigation bar is dark blue with the 'Era: Настройки' logo and a user profile icon. The sidebar on the left contains a list of configuration categories: Правила (vectorrules), Записи разговоров, Правила записи (recorders), Правила стенографирования (adv), Хранилища (storagrules), Сценарии, Голосовые меню (vnscripts), Служебные (svcsrpts), Задачи (svctasks), and Прочее. The main content area is titled 'Голосовые меню (/uc/vnscripts)' and features a table with the following columns: Наименование (name), Код для ссылки (code), and Примечания (opts.comment). The table lists several configurations, with 'Голосовые меню' highlighted in blue. The configuration details for 'Голосовые меню' are: Name: Голосовое меню, Code: out_02, Comment: Голосовое меню.

Наименование (name)	Код для ссылки (code)	Примечания (opts.comment)
Музыка 150 сек	in_01	
Правила записи (recorders)	in_02	
Правила стенографирования (adv)	in_03	
Хранилища (storagrules)	in_random	
Сценарии	out_01	
Голосовые меню (vnscripts)	out_02	
Служебные (svcsrpts)	out_03	
Задачи (svctasks)	out_04	
Прочее	out_05	
Голосовое меню	out_02	Голосовое меню

6.1.2. Эмулятор операторов

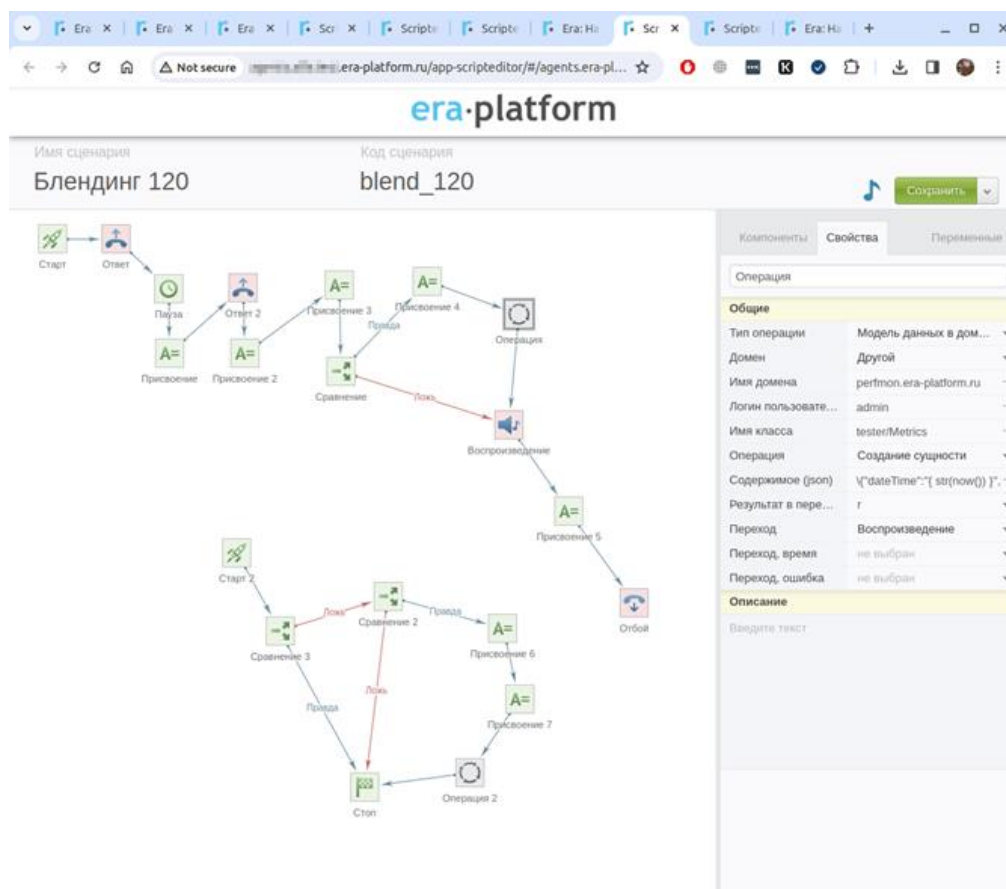
Поступающие на операторов вызовы обслуживаются системой-эмулятором операторов с помощью сценариев IVR по заданным в методике верификации показателям распределения.



Всего в системе-эмуляторе агентов заведено около 20 сценариев, отражающих различные сценарии из методики верификации:

Наименование (name)	Код для ссылки (code)	Примечание (opts.comment)
Блендинг 120	blend_120	
Блендинг слепой перевод	blend_blind	
Блендинг конференция	blend_conference	
Блендинг консультация и возврат	blend_consultation	
Блендинг удержание	blend_hold	
Блендинг random	blend_random	
Блендинг консультационный перевод	blend_transfer	
Консультация	consultation	
Консультация 30 сек	consultation30	
Входящие 120	in_120	
Входящие слепой перевод	in_blind	
Входящие конференция	in_conference	
Входящие консультация и возврат	in_consultation	
Входящие удержание	in_hold	
Входящие random	in_random	
Входящие консультационный перевод	in_transfer	
Музыка 30 секунд	ivr_music_30	
Исходящие 120	out_120	
Исходящие random	out_random	
Оператор второй линии	second_random	
Своя конференция	self_conf	

В каждом из этих сценариев производится замер и сохранение метрики. Например, в приведенном ниже сценарии производится замер времени обработки ответа на входящий вызов.

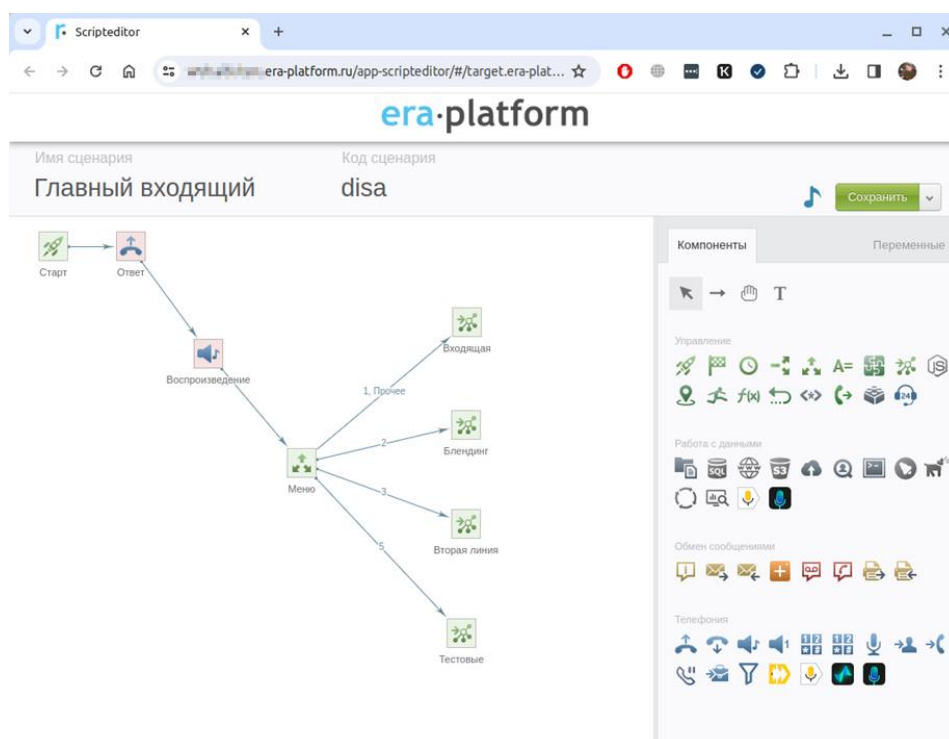


С учетом большой плотности вызовов сохранение в журнал показателей по метрикам производится выборочно с вероятностью от 1 до 100% в зависимости от частоты той или иной операции.

Данные по метрикам укладываются в БД и доступны на целевой системе в приложении Наблюдатель. В итоговой статистике отображаются общее число операций после вероятностного фильтра, максимальное и среднее значения.

6.1.3. Целевая система

Входящие вызовы распределяются в принимающим сценарием IVR целевой системы по различным очередям в соотношениях, установленных методикой верификации.



6.2. Сбор и анализ данных

Мониторинг и сбор данных осуществлялся в течение всего прохождения теста. После проведения теста проведена обработка и анализ результатов - выявление сбоев, блокировок, задержек, загруженность ресурсов.

6.2.1. Требования по утилизации аппаратных ресурсов

Ресурс	Описание	Значение
CPU (процессор)	Утилизация процессора в процентах от доступной	Не более 70%
RAM (память)	Утилизация оперативной памяти в процентах от доступной	Не более 80%
Disk (жесткий диск)	Занятость дискового пространства в процентах от доступной	Не более 60%
Сеть	Контроль трафика, пропускная способность (Throughput), потеря пакетов (Packet Loss), задержки (Latency),	Потеря пакетов не более 1%

6.2.2. Отчетность контакт-центра

Для подтверждения корректности результатов теста используются стандартные отчеты по работе операторов контакт-центра и обработке звонков, данные которых должны быть сопоставимы со статистикой эмулятора клиента и эмулятора оператора, например, количество звонков, которые сгенерил эмулятор клиента, должно совпадать с количеством принятых входящих звонков в контакт-центре.

7. Обзор стенда

7.1. Характеристики стенда

Стенд развернут в Яндекс.Облаке на следующих ресурсах:

Конфигурация	EraClient1	EraClient2	EraSrv1		EraSrv2		EraSrv3		EraSrv4	
Платформа	Intel Ice Lake	Intel Ice Lake	Intel Lake	Ice	Intel Lake	Ice	Intel Lake	Ice	Intel Lake	Ice
vCPU	16	16	32		32		32		32	
Гарантированная доля	100%	100%	100%		100%		100%		100%	
Объём RAM	32	32	32		32		32		32	
Прерываемая VM	Нет	Нет	Нет		Нет		Нет		Нет	
публичный IP-адрес	Нет	Нет	Да		Нет		Да		Нет	
Диск - 1 (загрузочный)	SSD 50	SSD 50	SSD 5		SSD 5		SSD 5		SSD 5	
Диск - 2			SSD 512		SSD 512		SSD 512		SSD 512	
Диск -3			SSD 512		SSD 512					

7.1.1. Целевая система

Единая четырехсерверная целевая система со следующим распределением ключевых ролей между серверами:

EraSrv1:

- web-сервер и бизнес-логика колл-центра,
- mg/rtx – медиашлюзы;

EraSrv2:

- dms - подсистема обработки данных и онлайн-уведомлений,
- mnesia - оперативная СУБД,
- mg/rtx – медиашлюзы;

EraSrv3:

- b2b - back-to-back SIP user agent,
- mware - передача событий со слоя телефонии в бизнес-логику,
- callstore - хранилище информации о текущих звонках
- mg/rtx - медиашлюзы,
- mix - микшер записанных разговоров из RTP в MP3;

EraSrv4:

- esg - стыковка по SIP с провайдерами,
- sg - стыковка по SIP с агентами,
- ivr - исполнение сценариев,
- mix - микшер записанных разговоров из RTP в MP3.

7.1.2.Эмуляторы

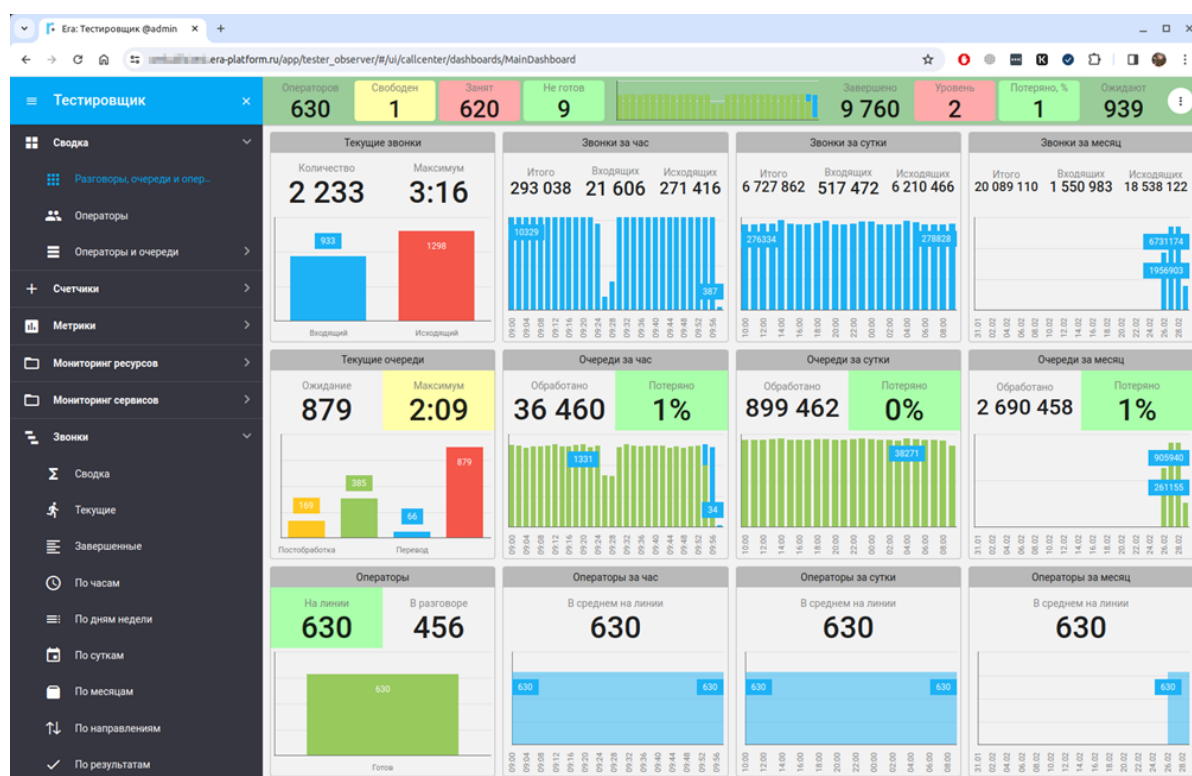
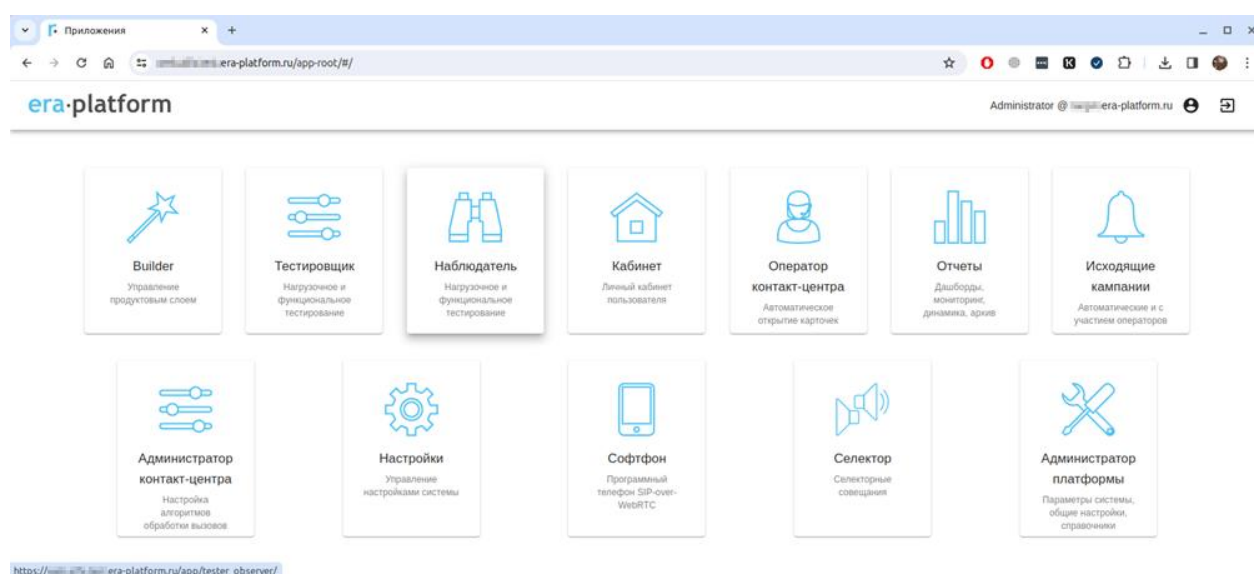
EraClient1 - эмулятор PSTN для приема трафика исходящей кампании.

EraClient2 - эмулятор PSTN для генерации трафика входящих звонков и операторских телефонов. На этой же машине запускается периодическое выполнение API-запроса для совершения входящего звонка, а также работают эмуляторы агентских рабочих мест для выставления через API результатов исходящих вызовов.

Функции СУБД распределены по этим же серверам для снижения общей нагрузки. Развернуто несколько экземпляров базы данных.

7.2. Приложение мониторинга

Для управления, мониторинга и верификации создано специальное веб-приложение “Наблюдатель”, включающее в себя одновременно и отчетность супервизора, и мониторинг метрик, и мониторинг утилизации аппаратных ресурсов.



7.3. Подключение тестовых телефонов

Выделено 10 учетных записей для подключения телефонных устройств по протоколу SIP. Пример настроек, выполненных для приложения (softphone) MicroSIP, для подключения к платформе представителей Заказчика представлен на скриншоте:

Аккаунт

Название аккаунта: Era_id1050004

SIP сервер:

SIP прокси: xxx.era-platform.ru:9160

Имя пользователя: 1050004

Домен: yyy.era-platform.ru

Логин: 1050004

Пароль: *****

Ваше имя:

Голосовая почта:

Префикс набора:

План набора:

☐ Hide Caller ID

Шифрование медиа: Отключено

Транспорт: UDP

Публичный адрес: Авто

Обновление регис...: 300 Keep-Alive: 15

☐ Публиковать присутствие

☐ Разрешить перезапись IP

☐ JSE

☐ Отключить таймеры сессии

Сохранить Отмена

- domain: yyy.era-platform.ru,
- outbound proxy: xxx.era-platform.ru,
- port: 9160,
- login, username, number, password: 1050001-1050005.

С тестовых телефонов 105XXXX можно звонить:

- 901XXXXXXX - выход в город с запуском randomного сценария входящих клиентов,
- 902XXXXXXX - выход в город с запуском randomного сценария исходящих клиентов с контактно-стью 5-10%,
- 9029999999 - выход в город с запуском randomного сценария исходящих клиентов с контактно-стью 100%,
- 903XXXXXXX - выход в город на тестовые линии (при наличии активных регистраций),
- 1XXXXXXX - прямой вызов любого оператора,
- 700 - главный входящий с донабором,
- 701 - очередь входящие,
- 702 - очередь входящие с блендингом,
- 703 - очередь исходящие,
- 704 - очередь вторая линия,
- 705 - очередь тестовые операторы,
- 799 - вечная музыка (отбой через два часа).

8. Проведение нагрузочного теста

26.02.2024 в 10:00 в ходе совместной ВКС стенд был проверен, совершен вызов с тестового аппарата. Произведенный вызов был обнаружен во всех мониторинговых отчетах и метриках, по нему была прослушана запись.

26.02.2024 в 10:08 был произведен запуск нагрузочного теста. Для этого был активирован процесс совершения вызовов на системе-эмуляторе PSTN (входящие вызовы плотностью 6 в секунду), процесс исходящей кампании на целевой системе (исходящие вызовы плотностью 80 в секунду), подключение 620 операторов, некоторые из которых, в соответствии с заданными в методике показателями, осуществляют login и logoff для замера времени обработки.

В ходе проведения теста:

1. Было совершено 18 565 353 исходящих вызовов за 72 часа. Количество исходящих вызовов по уровню соответствует расчетному. (расчетное – примерно $18\,750\,000 \text{ шт} = 72 \text{ часа} * 3600 \text{ секунд} * 80 \text{ вызовов в секунду} * 30/33$ (коэффициент времени остановки кампании каждые 30 минут) и поправка еще на 10 секунд).
Несущественное расхождение расчетных и фактических значений вероятно вызвано неточностью работы таймера инициации исходящих наборов в исходящих кампаниях – большинство по 80 в секунду, но иногда 79.
2. Количество обработанных входящих вызовов – 1 555 083 шт. (расчетное количество входящих вызовов: $1555200 = 72 \text{ часа} * 3600 \text{ секунд} * 6 \text{ вызовов в секунду}$)

29.02.2024 в 10:06 тест остановлен

9. Метрики и анализ технических результатов

9.1. Утилизация ресурсов

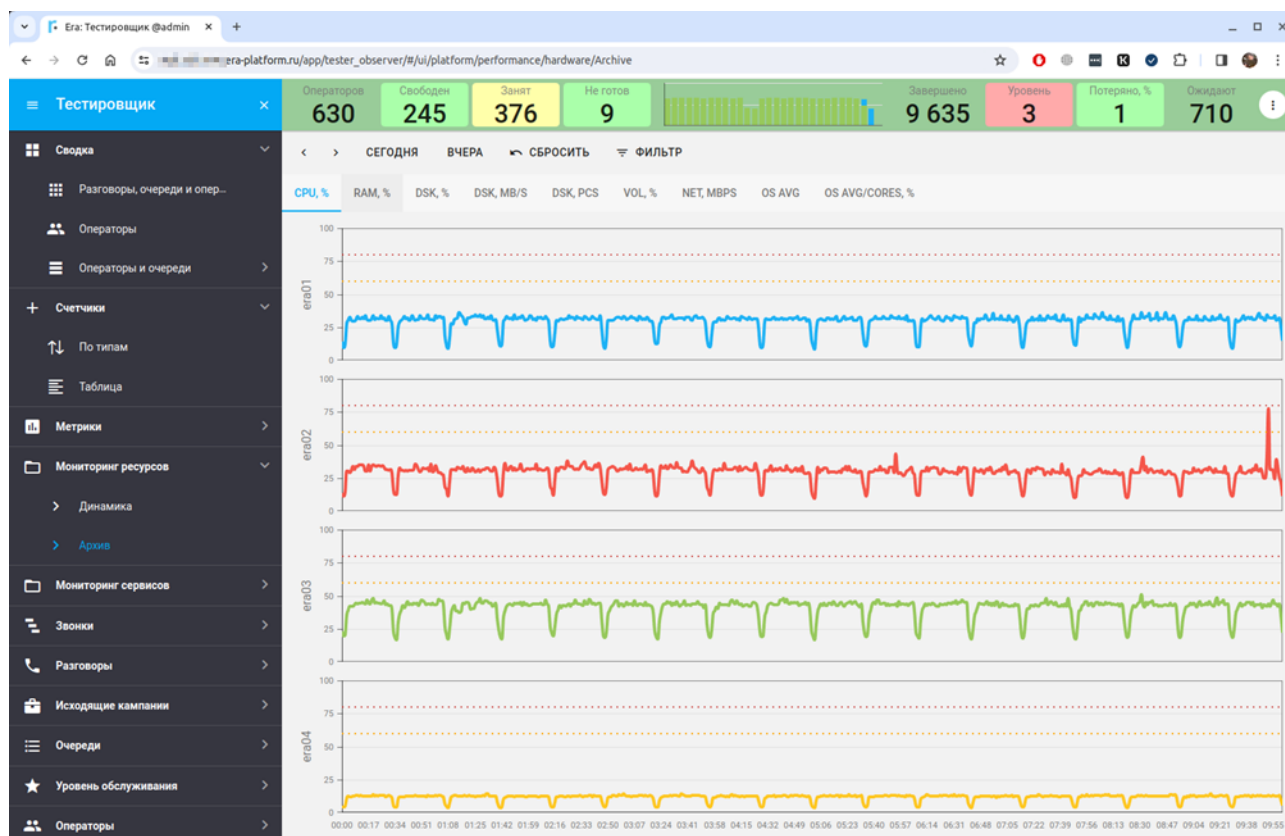
Приведенные ниже скриншоты сняты до остановки тестового стенда в случайный момент времени. На каждом скриншоте приведены соответствующие показатели по четырём серверам целевой системы.

Показатели снимаются отдельным микросервисом с помощью стандартных утилит операционной системы Linux (iostat, iotop, top, и т.д.) и укладываются в СУБД для возможности дальнейшего обращения.

Все показатели на скриншотах приведены за последние 4 часа и за последние сутки.

9.1.1. Утилизация CPU (процессоров)

Это общая загрузка процессора всеми активными процессами, исполняемыми на сервере, включая PostgreSQL.

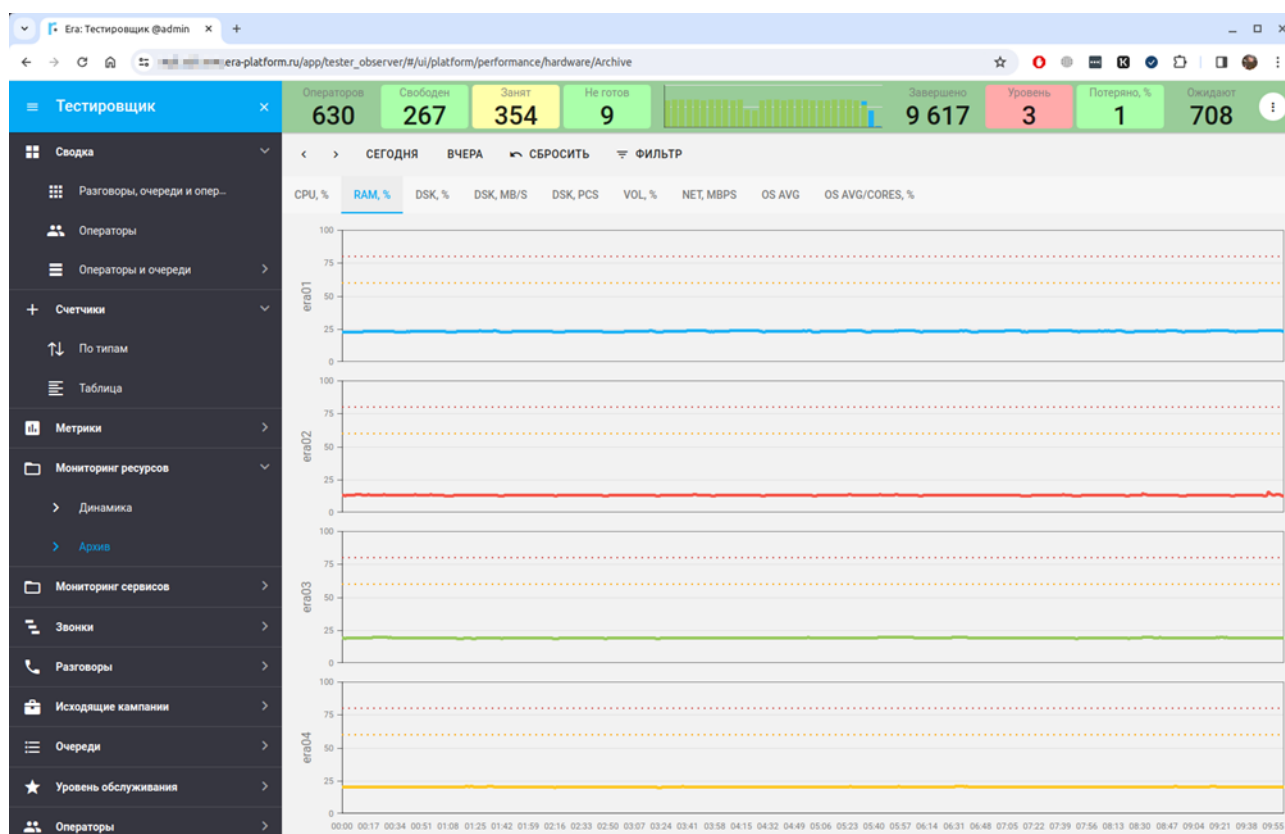
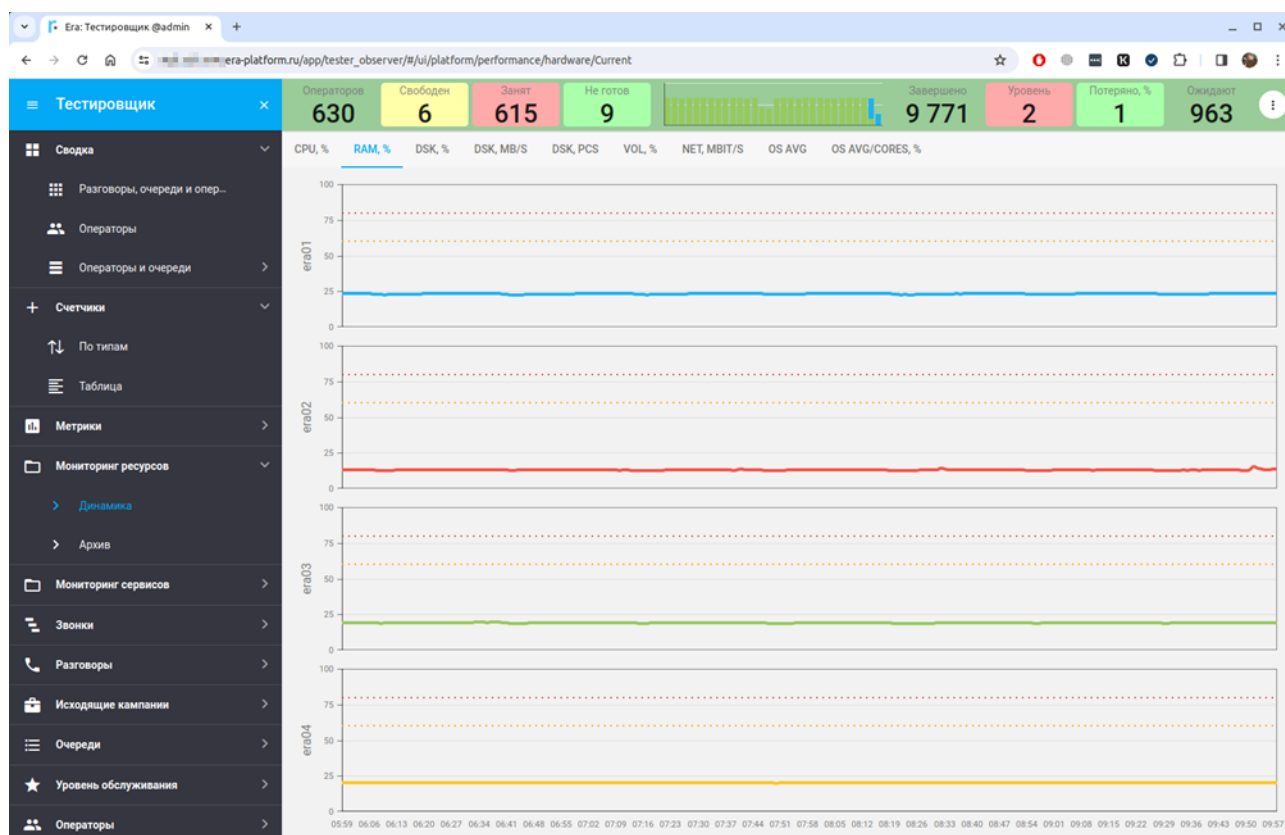


Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду. Точечный всплеск на втором сервере связан с обращениями к отчетности в СУБД PostgreSQL.

На протяжении теста никаких отклонений от целевой картины не наблюдалось.

9.1.2. Утилизация оперативной памяти

Это аллокация оперативной памяти всеми процессами операционной системы, выполняемыми на сервере, включая PostgreSQL.



На протяжении всего времени теста не наблюдается никаких заметных отклонений от константы, требуемой для обслуживания предусмотренных методикой количества и плотности вызовов и количества операторов.

Утечки памяти отсутствовали.

9.1.3. Утилизация жестких дисков

Показатель снимается регулярно утилитой iostat.

Это утилизация жестких дисков всеми процессами операционной системы, исполняемыми на сервере, включая PostgreSQL.





Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

Высокая нагрузка на первых двух серверах связана с размещением на них СУБД PostgreSQL. Два диска на первом сервере и два диска на втором сервере – отданы под 4 экземпляра PostgreSQL, каждый из них обслуживает по одной нагруженной таблице: таблица контрагентов для исходящей кампании, архивы звонков, архивы сеансов, архивы очередей.

9.1.4. Объем обмена данными с жесткими дисками

Показатель снимается регулярно утилитой iostat.

Это утилизация жестких дисков всеми процессами операционной системы, исполняемыми на сервере.



Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

Несмотря на плотную работу PostgreSQL с диском, объем данных, которые он фактически размещает на диск – минимален.

Система ведет запись и последующее микширование – это создает объем обмена данными с диском.

Система ведет логирование.

Система использует распределенную объектную БД, которая также проявляется в графиках (брокеры событий, хранилища текущих вызовов, и т.д.

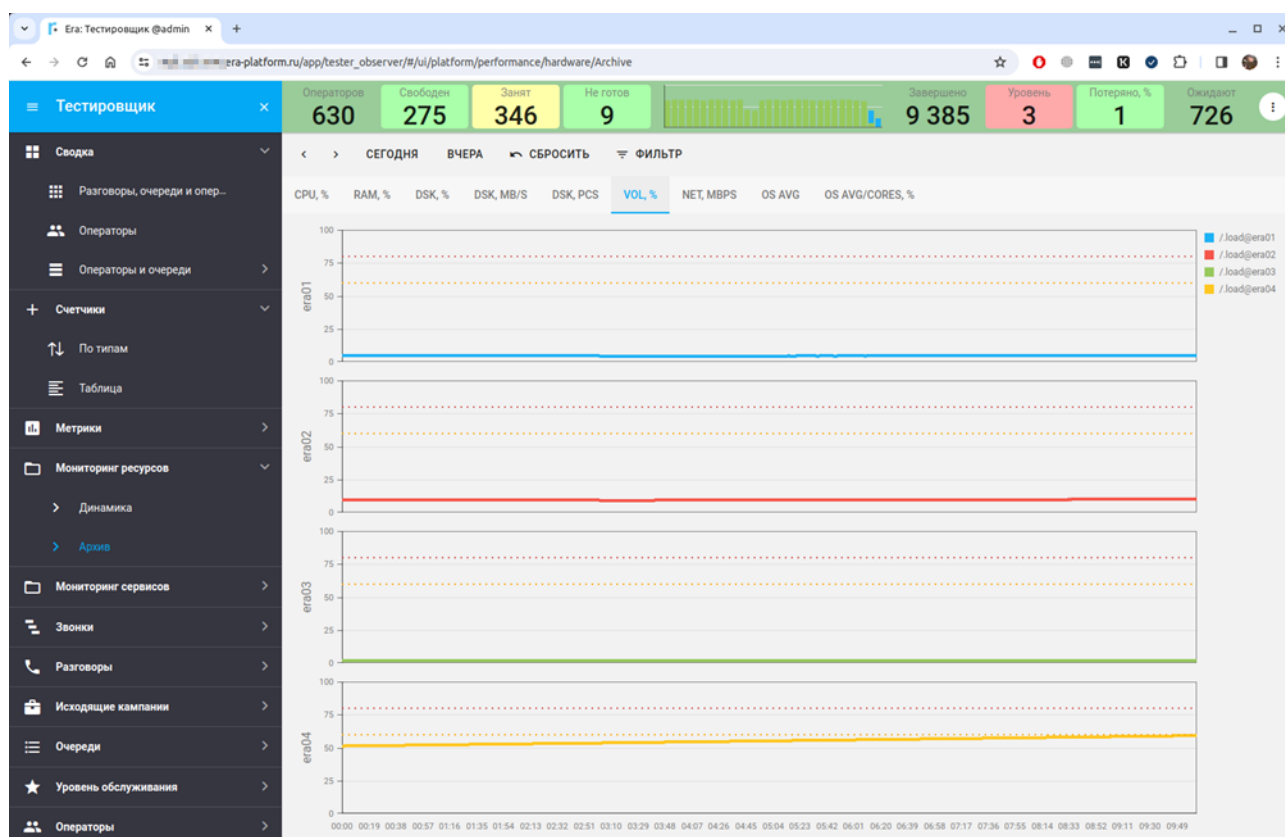
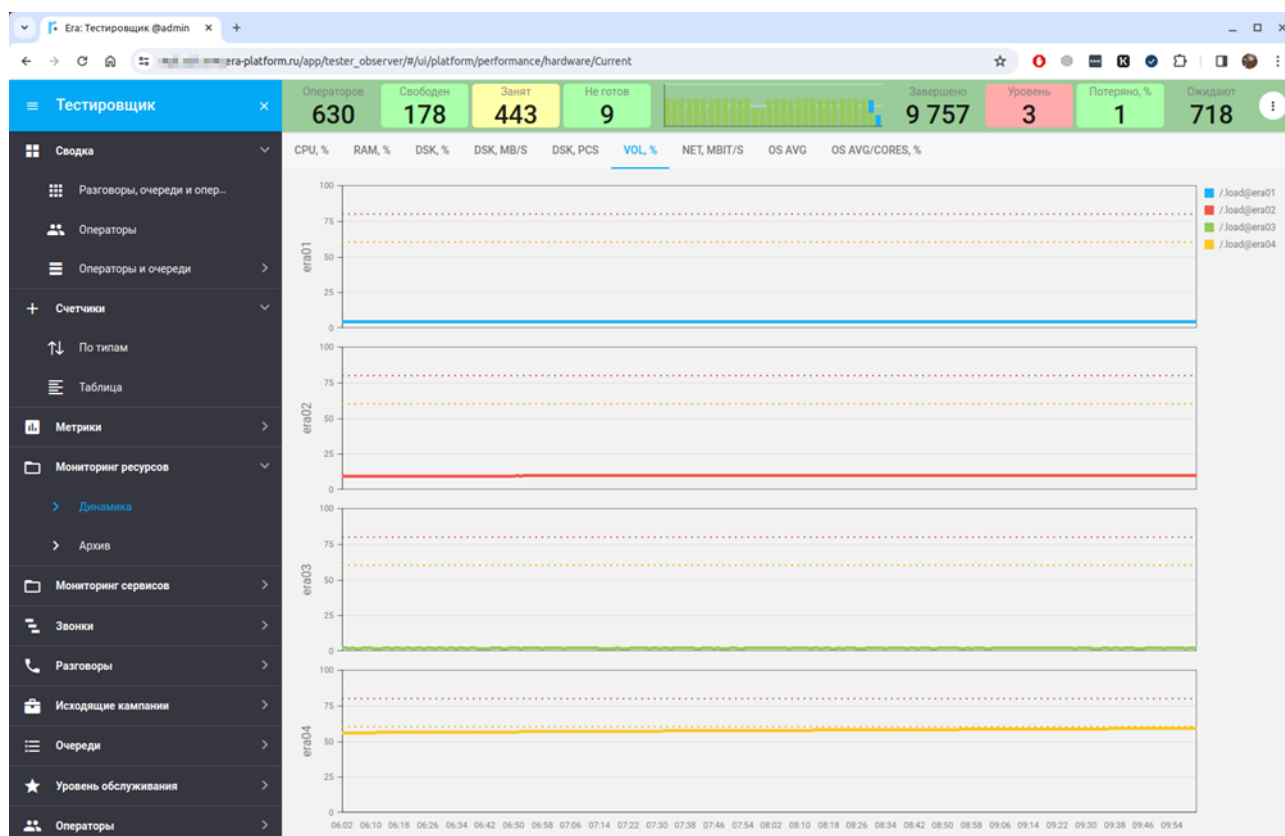
9.1.5. Количество операций к жестким дискам

Показатель снимается регулярно утилитой iostat.



Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

9.1.6. Объем занятого места на диске



Заметный плавный рост наблюдается только на 4-м сервере, куда складировались записанные разговоры.

9.1.7. Сетевой трафик



Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

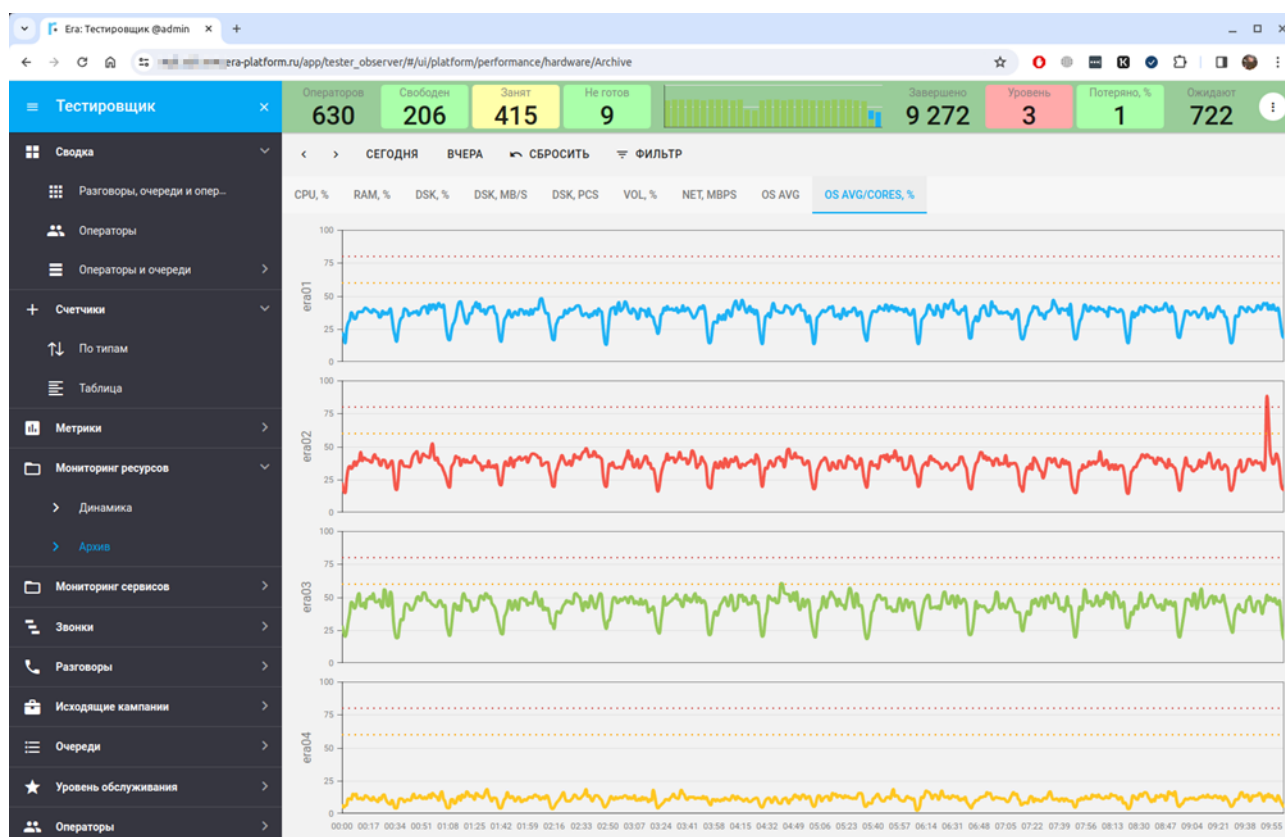
В большей степени выделяются серверы 1, 2 и 3, где размещены медиашлюзы. Свой вклад в сетевое взаимодействие вносят и SIP-сигнализация, и взаимодействие микросервисов.

9.1.8. AVG (показатель загрузки операционной системы Linux)

За последние 4 часа:

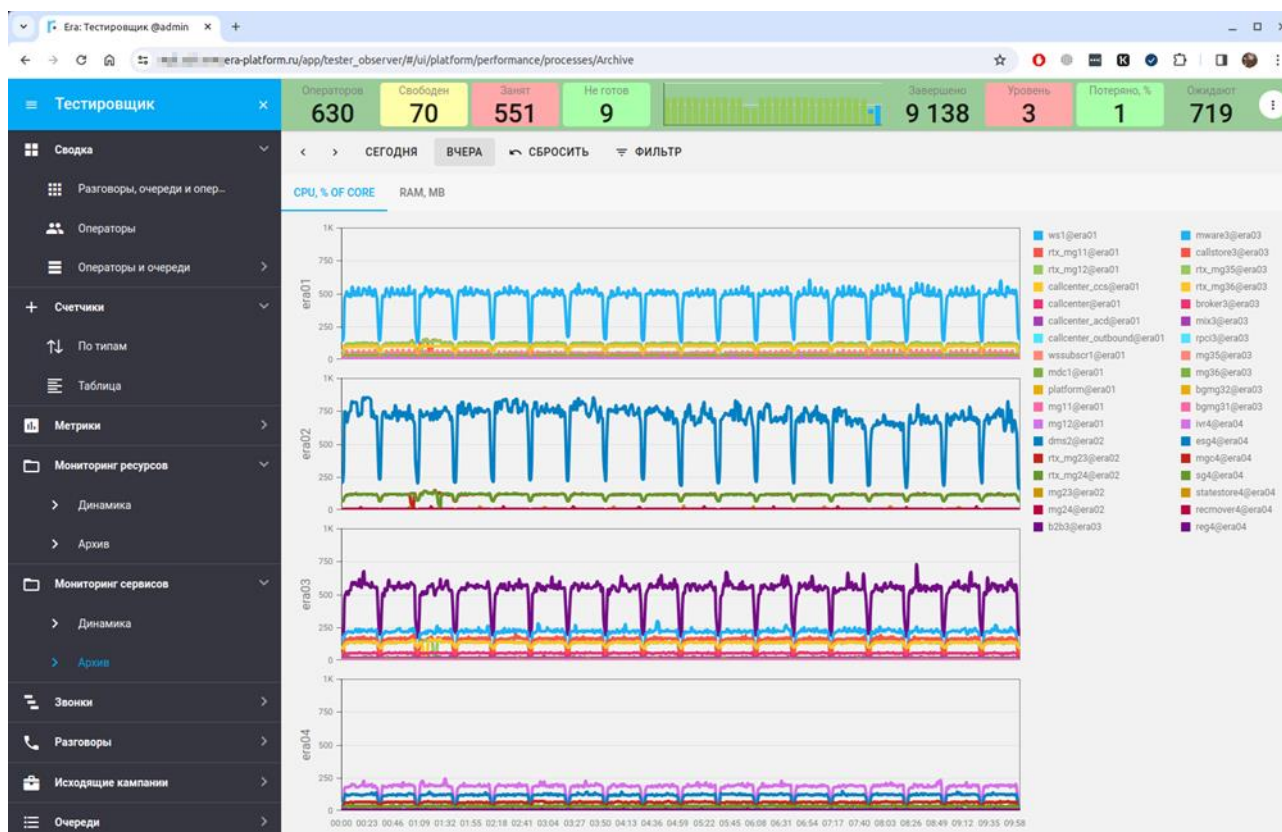


За последние сутки:



Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

9.1.9. Утилизация CPU микросервисами коммуникационной платформы



Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

Наиболее нагруженными микросервисами являются: на 1-м сервере – web-сервер, на 2-м сервере – менеджер модели данных, на 3-м сервере – back-to-back user-agent. Нагрузка по обслуживанию трафика распределена между 6-ю медиашлюзами на трех машинах. Микросервисы, занимающие менее 10% одного ядра CPU – в списке не отображаются, но могут быть включены в фильтре.

9.1.10. Утилизация оперативной памяти микросервисами коммуникационной платформы





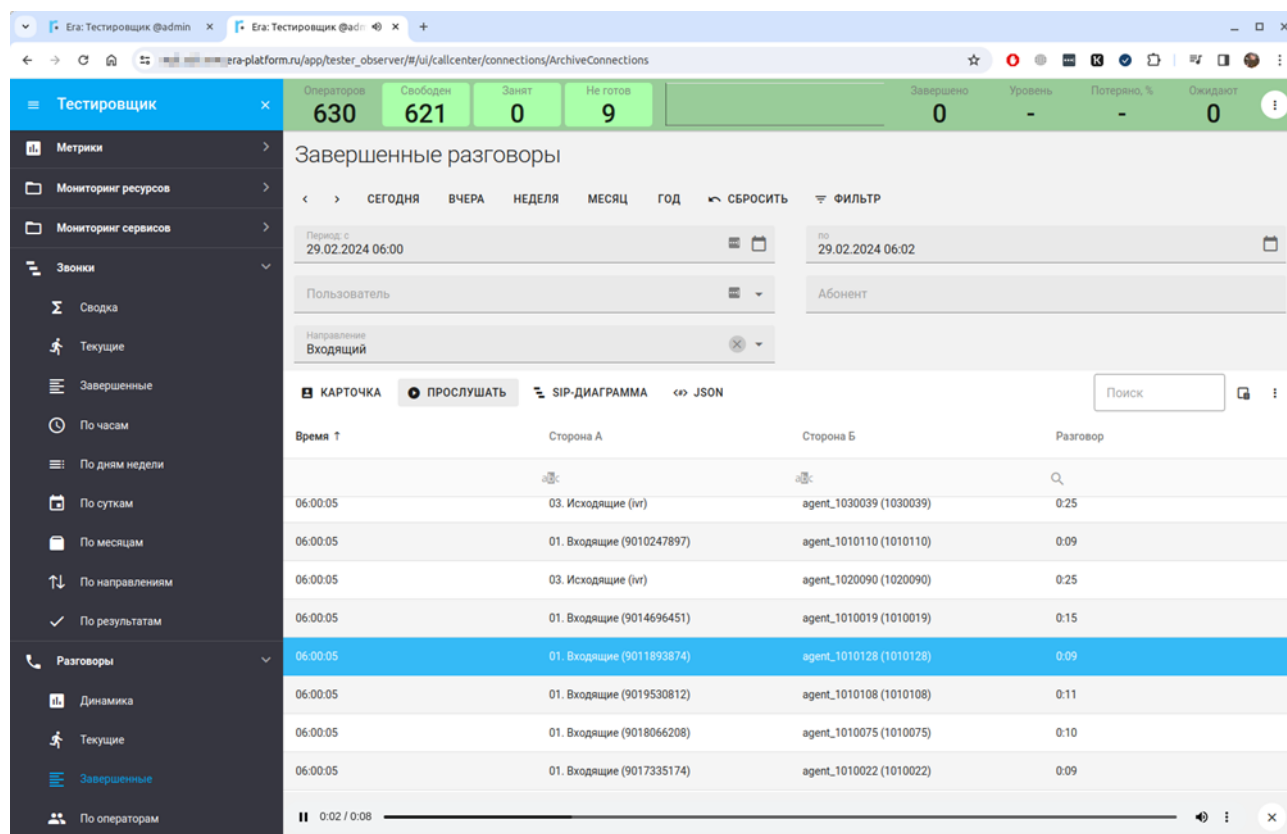
Наблюдаются характерные получасовые периоды интенсивной работы системы, связанные с активностью исходящей кампании плотностью 80 вызовов в секунду.

Кроме того, первый сервер проявляет пилу (зеленый цвет) – это загрузка исходящей кампанией списка контрагентов из БД по мере необходимости. При остановке кампании память высвобождается. Никаких растущих трендов не наблюдается.

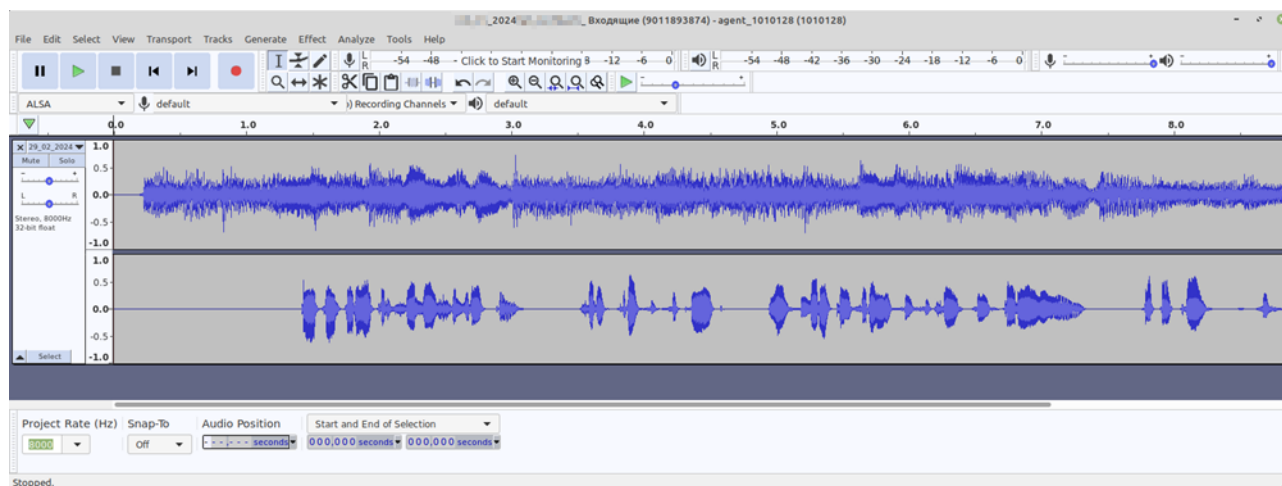
На трех серверах архивного графика за последние сутки видны распределенные во времени отсечки вниз – это плановый ежесуточный мягкий (gentle) вывод медиашлюзов из обслуживания.

9.2. Записи разговоров

В левом канале записан трафик от эмулятора PSTN – мелодия. В правом канале записан трафик от эмулятора агента – синтезированная речь (стихи).



На скриншоте приведен пример записанного mp3-файла, открытый в редакторе Audacity.



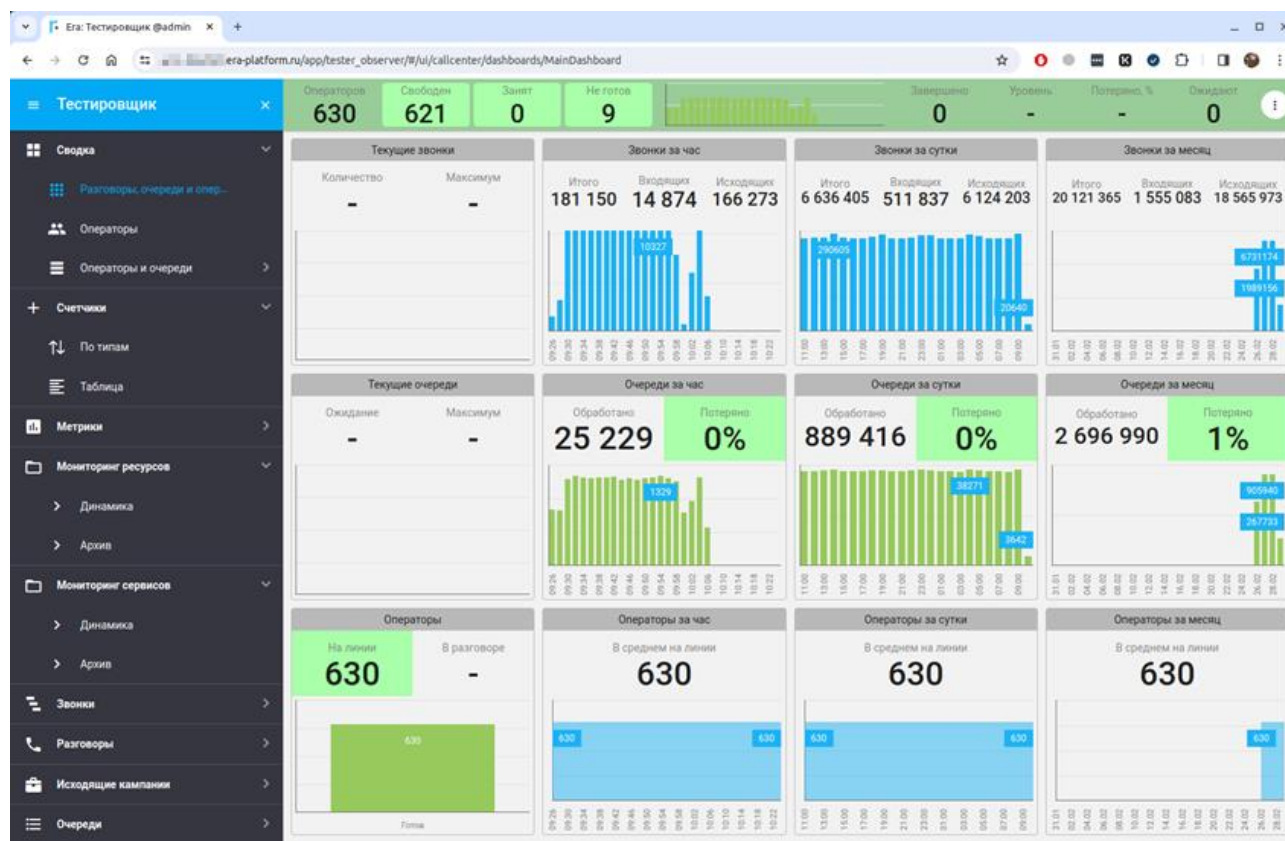
После завершения теста также был проведён выборочный анализ записанных разговоров. Пропушаны звонки с проверкой качества записи.

9.3. Остановка теста

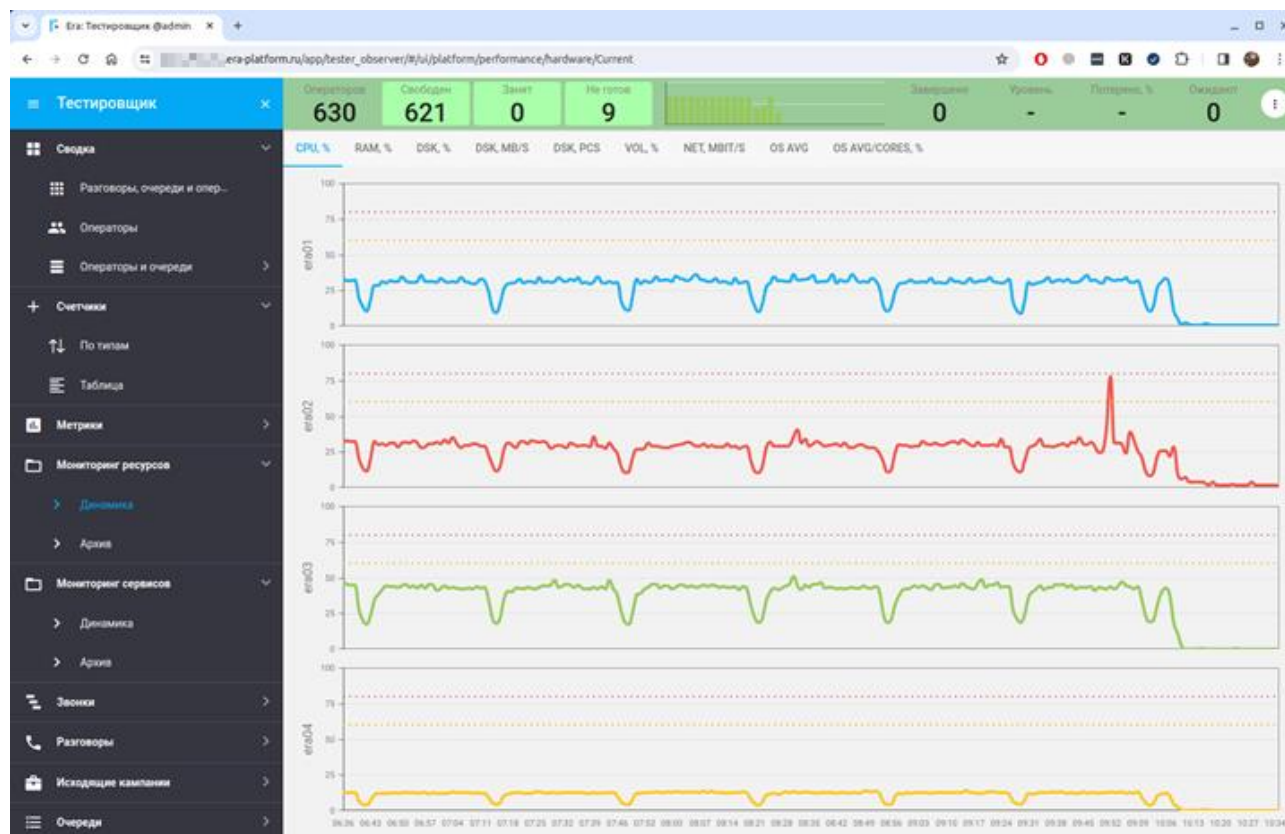
Тест остановлен 29.02.2024 в 10:06.

Спустя 3 минуты после остановки 620 операторов освободились от вызовов и от поствызывной обработки.

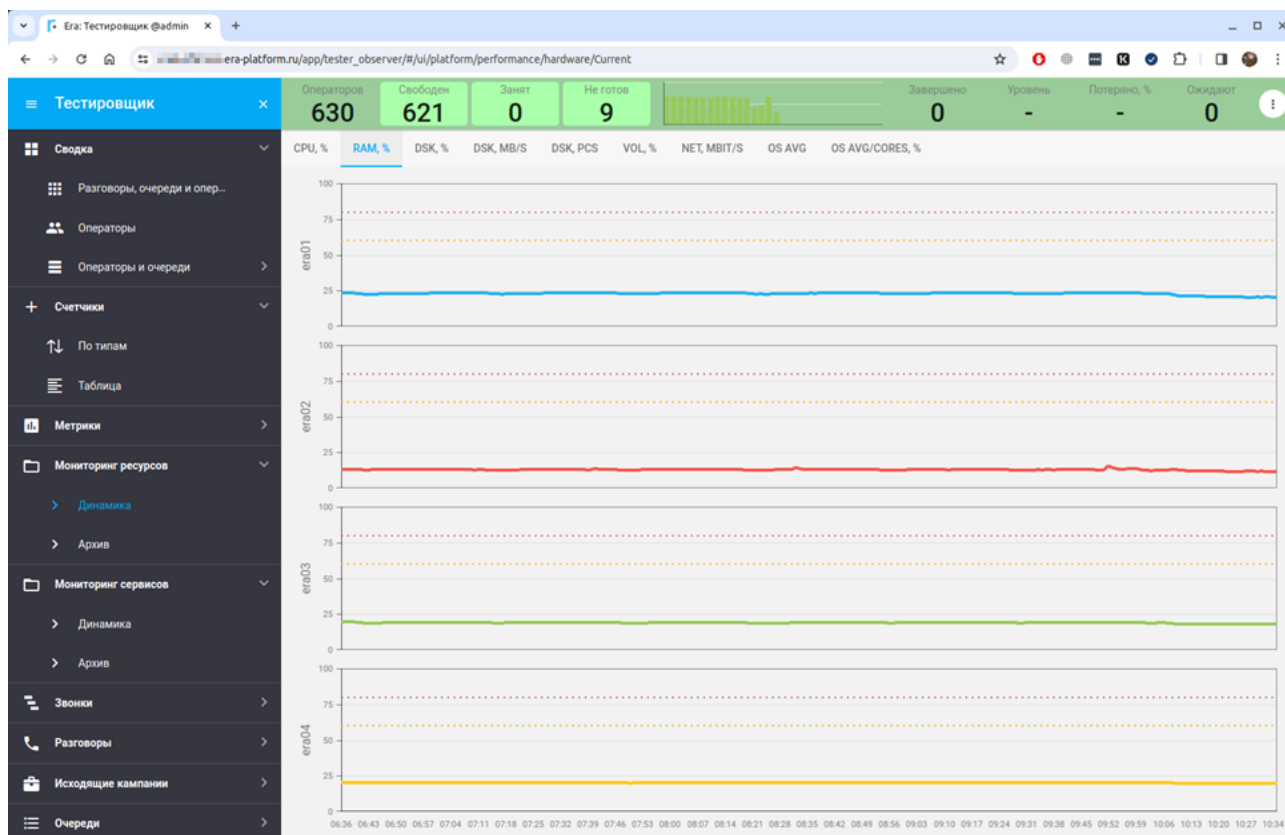
Система высвободила все ресурсы, требуемые для обслуживания интенсивного потока вызовов, и перешла в idle-состояние.



Утилизация CPU:



Утилизация оперативной памяти:



Утилизация жестких дисков:



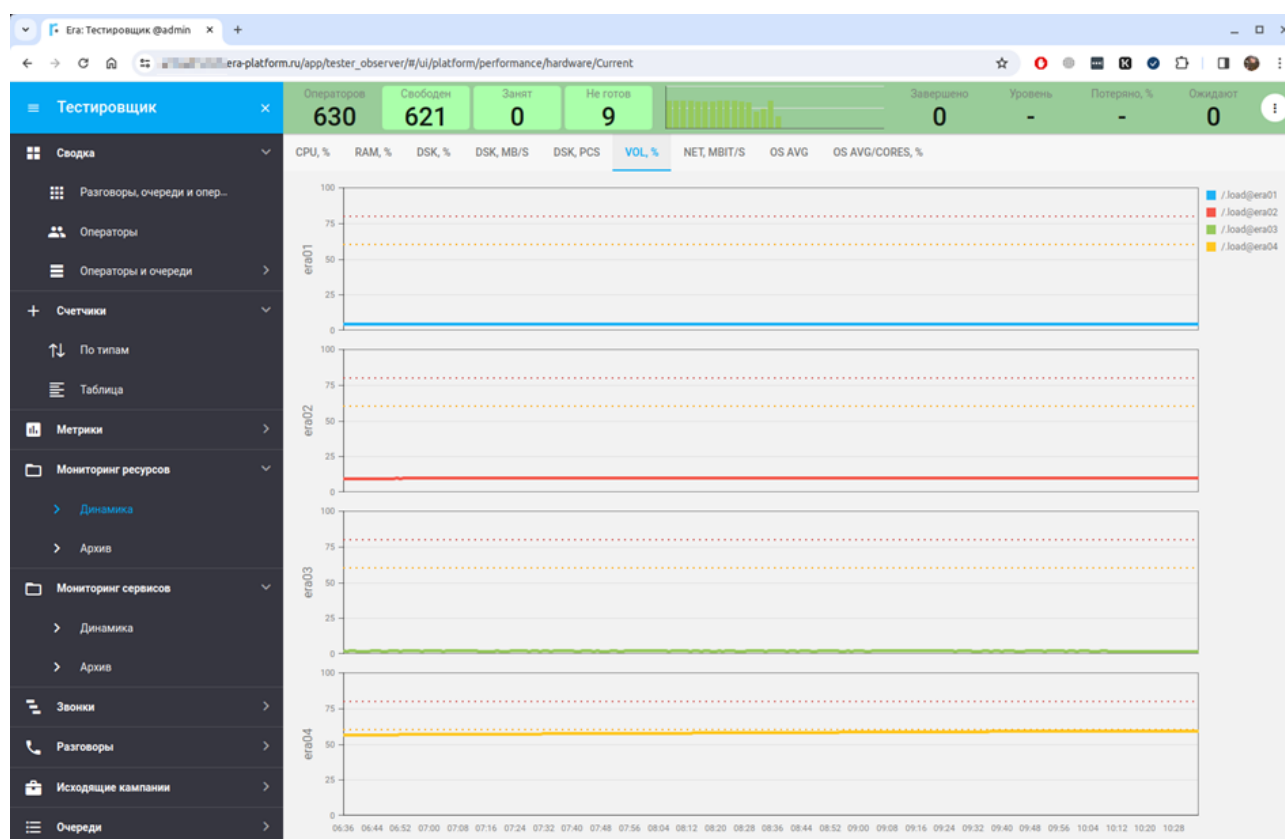
Объем данных, обмениваемых с жесткими дисками:



Количество операций с жестким диском:



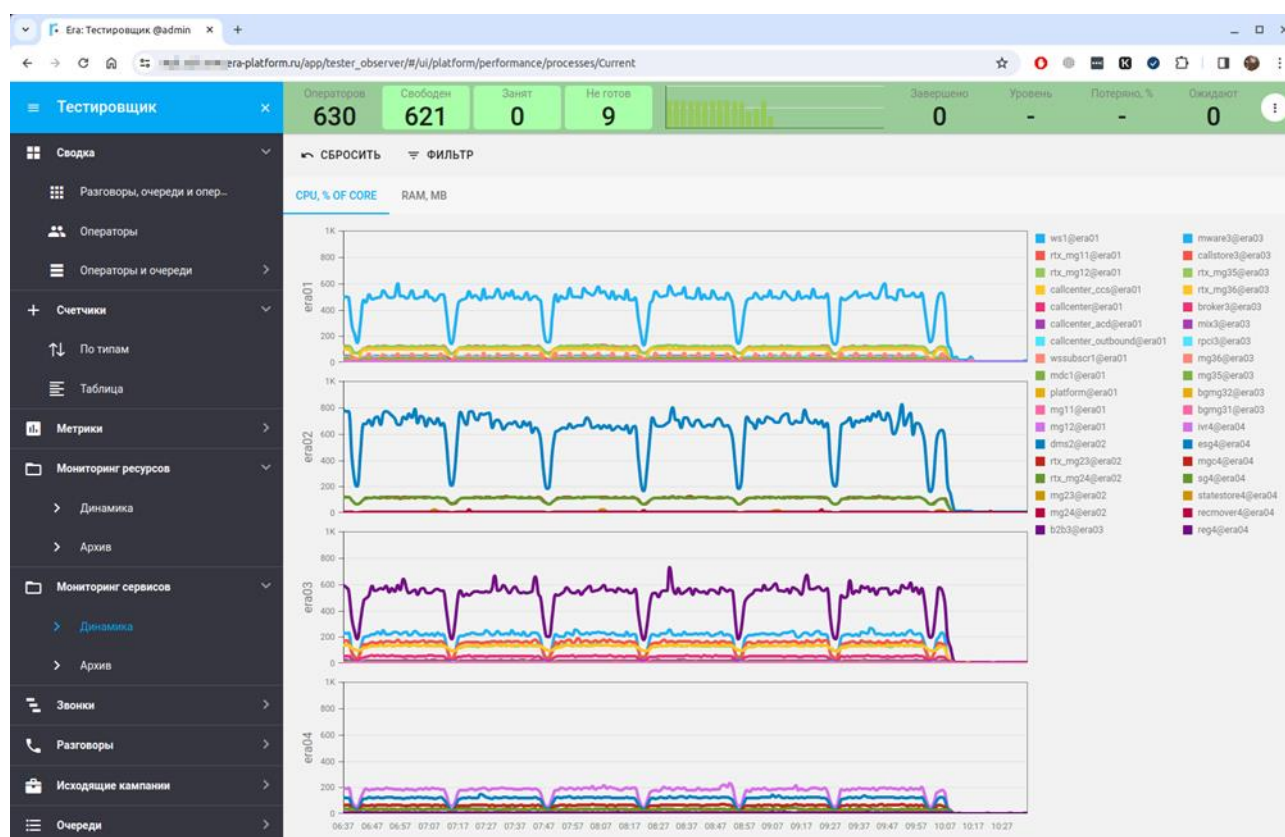
Свободное место на жестких дисках:



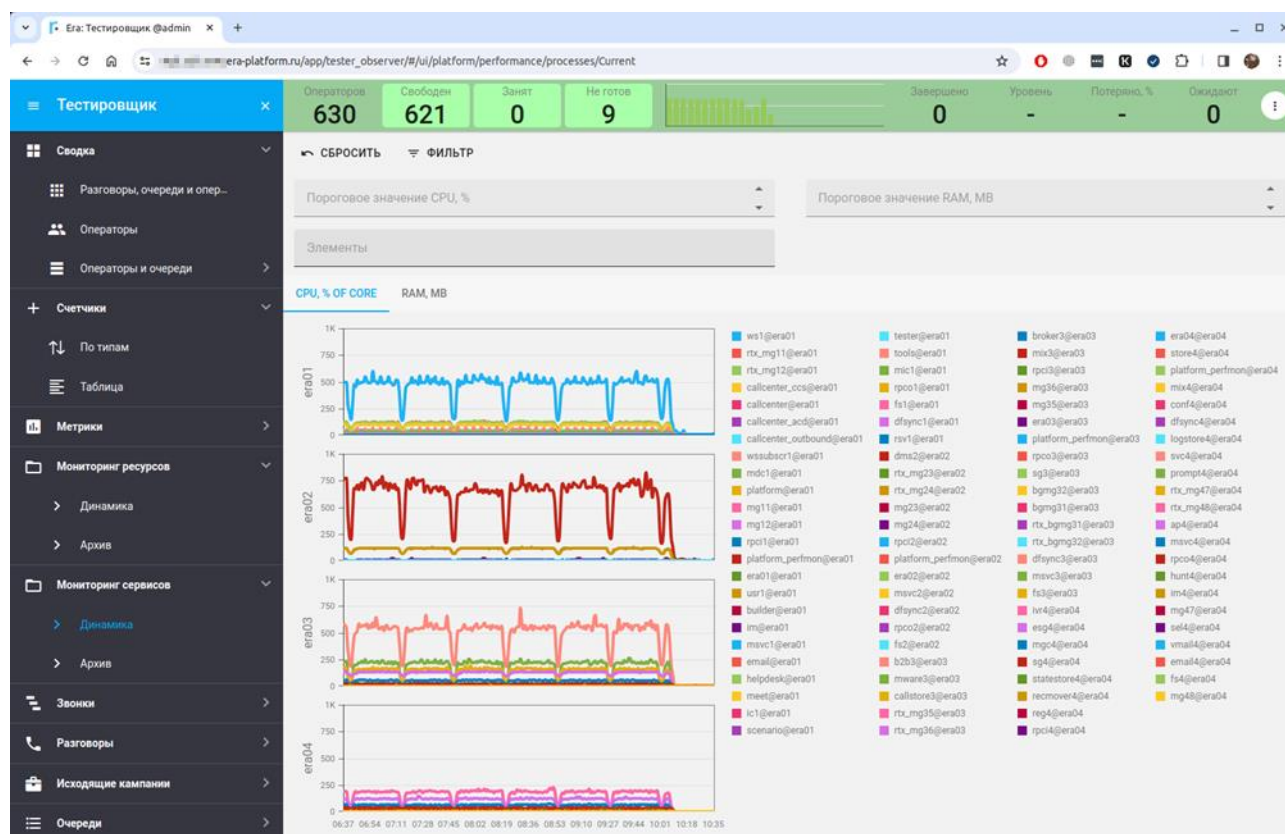
Объем данных, передаваемых по сети:



Утилизация CPU наиболее нагруженными микросервисами коммуникационной платформы:



Утилизация CPU всеми микросервисами коммуникационной платформы:



Эра: Тестировщик @admin

era-platform.ru/app/tester_observer/#/u/platform/performance/processes/Current

Операторов: 630 Свободен: 621 Занят: 0 Не готов: 9

Завершено: 0 Уровень: - Потеряно, %: - Ожидает: 0

Сводка

Разговоры, очереди и опер...

Операторы

Операторы и очереди

Счетчики

По типам

Таблица

Метрики

Мониторинг ресурсов

Динамика

Архив

Мониторинг сервисов

Динамика

Архив

Звонки

Разговоры

Исходящие кампании

Очереди

СБРОСИТЬ ФИЛЬТР

CPU, % OF CORE RAM, MB

era01

era02

era03

era04

ws1@era01

wssubscr1@era01

callcenter_outbound@era01

callcenter_ccs@era01

rtx_mg11@era01

rtx_mg12@era01

dms2@era02

b2b3@era03

mware3@era03

callstore3@era03

ivr4@era04

esp4@era04

reg4@era04

06:37 06:45 06:53 07:01 07:09 07:17 07:25 07:33 07:41 07:49 07:57 08:05 08:13 08:21 08:29 08:37 08:45 08:53 09:01 09:09 09:17 09:25 09:33 09:41 09:49 09:57 10:05 10:13 10:21 10:29

The screenshot shows the 'Тестировщик' (Tester) application interface. The top bar displays statistics: 630 Operators, 621 Free, 0 Occupied, 9 Not ready. Below are tabs for 'Сводка' (Summary), 'Разговоры, очереди и операторы' (Conversations, queues and operators), 'Операторы' (Operators), 'Операторы и очереди' (Operators and queues), 'Счетчики' (Counters), 'По типам' (By type), and 'Таблица' (Table). The 'Метрики' (Metrics) section is active, showing 'Мониторинг ресурсов' (Resource monitoring) and 'Мониторинг сервисов' (Service monitoring). The main area displays four line charts for CPU % of CORE and RAM, MB usage across four eras (era01, era02, era03, era04). A legend on the right lists various services and their corresponding colors.

9.4. Счетчики

Тестировщик				Операторов	Свободен	Занят	Не готов	Завершено	Уровень	Потеряно, %
				630	621	0	9	0	-	-

Счетчики по типам

< > СЕГОДНЯ ВЧЕРА НЕДЕЛЯ МЕСЯЦ ГОД СБРОСИТЬ ФИЛЬТР ЭКСПОРТ

Имя ↑	Количество
call.agents	2702558
call.agents.blend	565666
call.agents.in	1525747
call.agents.out	528764
call.agents.second	82347
call.in	1554440
call.in.dtmf1	626847
call.in.dtmf2	141408
call.in.wait	786185
call.out	18565353
call.out.busy	151751
call.out.human180	455272
call.out.human60	611197
call.out.noanswer	152565
call.out.voicemail	137333

9.5. Метрики

Тестировщик				Операторов	Свободен	Занят	Не готов	Завершено	Уровень	Потеряно, %	Ожидает	Максимум
				630	621	0	9	0	-	-	0	-

Метрики по типам

< > СЕГОДНЯ ВЧЕРА НЕДЕЛЯ МЕСЯЦ ГОД СБРОСИТЬ ФИЛЬТР ЭКСПОРТ

Имя ↑	Количество	Среднее	Максимум	До 1 сек	До 2 сек	До 3 сек	До 1 сек, %	До 2 сек, %	До 3 сек, %
acd.call.place	2347	66.62	573	2347	2347	2347	100	100	100
acd.call.toFreeAgent	3637	8.81	25	3637	3637	3637	100	100	100
agent.logoff	1689	36.04	254	1689	1689	1689	100	100	100
agent.logon	1689	457.63	688	1689	1689	1689	100	100	100
call.answer	3201	30.55	91	3201	3201	3201	100	100	100
call.conference	1920	1201.04	1549	0	1920	1920	0	100	100
call.drop	2840	1.88	4	2840	2840	2840	100	100	100
call.make	2605	178.77	869	2605	2605	2605	100	100	100
call.transfer.blind	3583	125.83	317	3583	3583	3583	100	100	100
call.transfer.finish	3128	181.4	627	3128	3128	3128	100	100	100
call.transfer.start	2098	175.77	586	2098	2098	2098	100	100	100
rtp.delay.A.ms	3329	116.29	439	3329	3329	3329	100	100	100
rtp.delay.B.ms	3329	116.63	439	3329	3329	3329	100	100	100
rtp.lost.A.%	3329	0	0.23						
rtp.lost.B.%	3329	0	0.15						

Для расчета метрик использовалась таблица в PostgreSQL, куда данные заносились сценариями, исполняемыми на целевой системе и системе-эмуляторе агентов. Ввиду высокой нагрузки трафика в целях ослабления давления на БД, для расчета и сохранения принимались не все

вызовы, а часть. Выборка производилась случайно исходя из соотношения вероятностей, чтобы общее количество операций за 72 часа тестирования было более 10 тысяч, но не слишком большим.

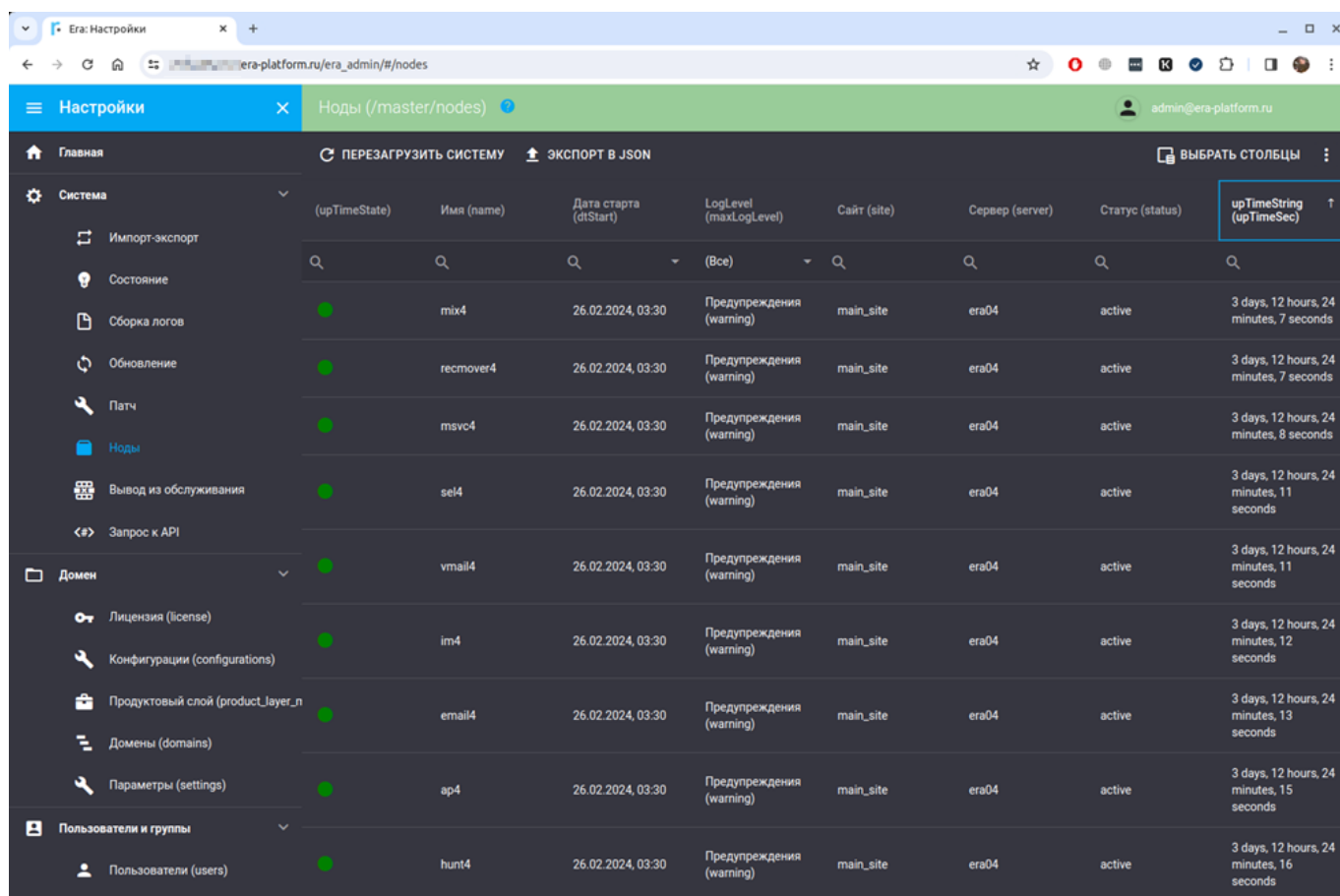
Процент потерь в голосовом трафике (качество передачи голоса) в среднем составил 0, максимальные потери составили 0.15%. Качество трафика остается высоким, несмотря на сложность сети в Яндекс.Облаке. Разнесение медиашлюзов по разным серверам позволило устранить проблему потерь в трафике, вызванную качеством сети и виртуального окружения.

Из всех показателей только время сбора конференции из трех абонентов вышло за пределы 1 секунды. Тем не менее в 10% случаях оставаясь в пределах установленного методикой верификации показателя в 2 секунды и 3 секунды.

Среди тяжелых операций – логин. В среднем занимает 450 мс, максимально был 1.5 секунды. Всего за пределы 1 секунды вышло 5 операций из 12 тысяч.

Прочие операции: инициация вызова, ответ на вызов, отбой, постановка на удержание, перевод и другие, собранные на основании методики верификации – выполнялись в среднем менее 200 мс.

Ни один микросервис за время теста не перезапускался:



(upTimeState)	Имя (name)	Дата старта (dtStart)	LogLevel (maxLogLevel)	Сайт (site)	Сервер (server)	Статус (status)	upTimeString (upTimeSec)
🟢	mix4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 7 seconds
🟢	rectmover4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 7 seconds
🟢	msvc4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 8 seconds
🟢	sef4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 11 seconds
🟢	vmal4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 11 seconds
🟢	im4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 12 seconds
🟢	email4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 13 seconds
🟢	ap4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 15 seconds
🟢	hunt4	26.02.2024, 03:30	Предупреждения (warning)	main_site	era04	active	3 days, 12 hours, 24 minutes, 16 seconds