

Как взять под контроль инфраструктурные расходы

[servercore]

СЕРГЕЙ БОБКО
Директор
по региональному
развитию



О чем доклад

Как построить процессы,
которые помогут
оптимизировать расходы
на инфраструктуру



Сценарии расходов на инфраструктуру

[В РЕАЛЬНОМ МИРЕ И РЕАЛЬНОМ БИЗНЕСЕ]

От сценария зависят потребности компании
и используемые инструменты.



СЦЕНАРИИ РАСХОДОВ НА ИНФРАСТРУКТУРУ

[ЕДИНОЕ УПРАВЛЕНИЕ ИНФРАСТРУКТУРОЙ]

Сценарии расходов на инфраструктуру

[ЕДИНОЕ УПРАВЛЕНИЕ ИНФРАСТРУКТУРОЙ]

- Характерны постоянные нагрузки
- Масштабирование происходит планоно, с согласованием
- Часто развертывают по модели on-premise

Забутые неиспользуемые и избыточные ресурсы увеличивают стоимость инфраструктуры

При чем здесь облачные сервисы

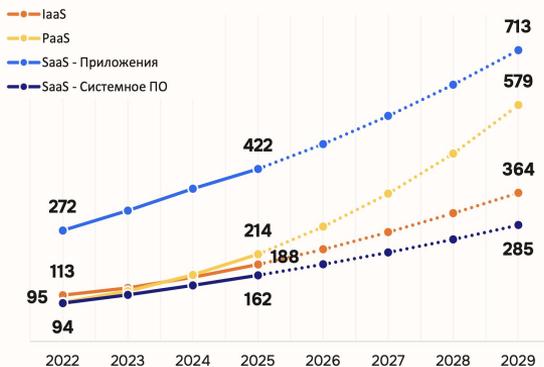
Рынок облачных сервисов стремительно развивается

Облачные сервисы упрощают доступ к ресурсам

В расходах появляются неожиданные koszty

Глобальный рынок публичных облачных сервисов в 2022–2029* в разрезе сегментов

USD млрд



Источник: IDC's Worldwide Software and Public Cloud Services Spending Guide

* Данные за 2026–2029 годы – прогноз

СЦЕНАРИИ РАСХОДОВ НА ИНФРАСТРУКТУРУ

[1] Автомасштабирование по потребности

[2] Децентрализованный заказ ресурсов

[1] Автомасштабирование по потребности

[ПЕРЕМЕННЫЕ НАГРУЗКИ]

- Сезонные и суточные пики (e-commerce, маркетинг)
- Временные окружения для разработки и тестирования
- Тяжелые расчеты и ML-модели (GPU-ресурсы)
- Быстрые MVP, PoC и выход в новые регионы

[1] Автомасштабирование по потребности

[ИНСТРУМЕНТЫ]

- Автоскейлинг в облачных сервисах и хранилищах
- Kubernetes и другие оркестраторы
- Terraform-провайдеры и сценарии управления

[1] Автомасштабирование по потребности

[РИСКИ]

- Появление «белых пятен» в расходах при быстром росте
- Недостаток прозрачности: сложно понять, какие ресурсы реально нужны

[2] Децентрализованный заказ ресурсов

[ПЕРЕМЕННЫЕ НАГРУЗКИ]

- Команды и проекты получают права на самостоятельный заказ ресурсов
- Подтверждение не требуется — инфраструктура создается через self-service

[2] Децентрализованный заказ ресурсов

[ПРЕИМУЩЕСТВА]

- Выше скорость — нет задержек на согласованиях
- Гибкость — каждая команда оптимизирует под свои задачи

[2] Децентрализованный заказ ресурсов

[РИСКИ]

- Нет единого контроля — у каждой команды свои правила
- Рост незапланированных расходов — разные приоритеты, дублирование ресурсов

ЦЕННОСТИ ОБСЕРВАБИЛИТИ

[1] Прозрачность затрат

[2] Возможность оптимизации затрат на инфраструктуру

[3] Объяснение бизнесу необходимости расходов

Обсервабилити и контроль

- Это процесс, а не разовое действие
- Хочешь управлять — измеряй расходы

[ЦЕЛЬ] Взять под контроль инфраструктурные расходы

Контроль в моменте

- Это значит, что вчерашние расходы нужно отслеживать уже сегодня, а не спустя время
- Чем короче период контроля, тем понятнее расходы

ХОРОШАЯ ПРАКТИКА

Контролировать расходы
вчерашнего дня сегодня

ПЛОХАЯ ПРАКТИКА

Контролировать расходы
прошлого месяца сегодня

За счет чего теряют деньги

—
Неиспользуемые
ресурсы

—
Сетевые
артефакты

—
Хранение данных
без политики

—
Автоскейлинг без контроля

—
Общие причины

КАКИЕ ИНСТРУМЕНТЫ
ПРЕДЛАГАЕТ КЛИЕНТАМ
[Servercore]

Далее рассмотрим

- Как контролировать расходы уже сегодня
- Как прогнозировать бюджет и на сколько дней его хватит
- Как достичь прозрачности инфраструктуры и внедрить процессы регулярной оптимизации

[ЗАДАЧА] Не сокращать косты на инфраструктуру, а извлечь максимальную бизнес-ценность из каждого тенге

Контроль в моменте

Transactions

All time



All types of funds



All activity



All products



Download CSV

February 2026

+KZT 400,000.00 / -KZT 405,505.24

23.02 00:50	Payment for service Managed Kubernetes	-KZT 538.87
23.02 00:32	Payment for service Object S3 Storage	-KZT 202.88
23.02 00:20	Payment for service Managed Databases	-KZT 1,943.63
23.02 00:02	Payment for service Cloud servers	-KZT 4,216.21
22.02 00:32	Payment for service Object S3 Storage	-KZT 540.84
22.02 00:30	Payment for service Managed Kubernetes	-KZT 1,556.75

Platform consumption

Current cost of infrastructure Consumption graphics External traffic

Current cost

KZT 19,178.40

Hour

Day

Month

Actual to 23.02.2026, 08:00. The next update - from 11:00 to 12:00.
Changes after resource creation or deletion may take up to 2 hours to appear, even after a scheduled update.

Detailization

By project



All projects



All resources



Project/metric	Used	Total
▼		KZT 339.12
▼		KZT 4,829.52
▼		KZT 5,691.84
▼		KZT 8,317.92

Platform consumption

Current cost of infrastructure [Consumption graphics](#) External traffic

Consumed

Paid ☺

01.11.2025



23.02.2026



Month



By project



All projects



All resources

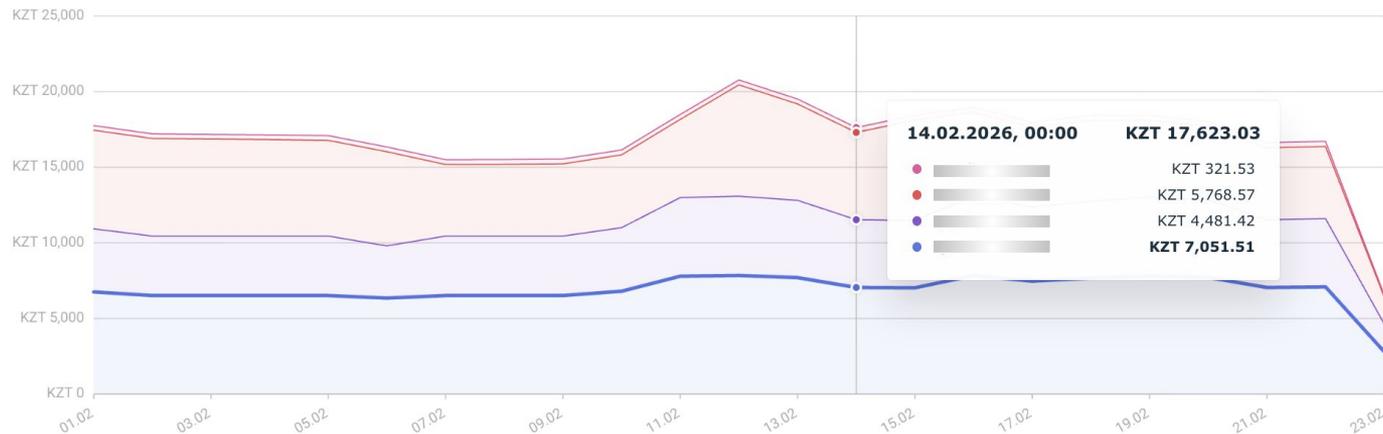


Details

Download CSV

1–23 February 2026

KZT 392,018.3



Что такое предсказуемость расходов

- Способность заранее оценивать горизонт и понимать, на какой период при текущем потреблении и будущих платежах хватит доступного баланса

Прогноз и уведомление

Billing overview

Balance

Main

KZT 305,169.04

Enough for ≈ 18 days

Balance notifications

When the balance falls below the threshold, we'll notify to the [«Balance and payments» category recipients](#).

Balance limit

80 000

KZT

Send emails every

3 hours



Остаток баланса

Остаток в днях

=

Среднее потребление за 3 часа

+

Предстоящие списания

Предсказуемость — это уверенность в завтрашнем дне

СИГНАЛ

Осталось N дней
+ праздники



РЕАКЦИЯ

- [1] Пополнить баланс
- [2] Проверить аномалии
- [3] Скорректировать потребление

Что такое прозрачность

- Состояние, при котором стейкхолдерам понятно, кто на что потратил деньги
- Если все ресурсы лежат в одном «котле», то никто не отвечает за эти траты

Что такое проекты

- Логическая группа ресурсов:
 - VM,
 - Диски
 - балансеры и другое
- Единица бюджета, владельца, отчетности

[ПРОЕКТЫ] — это границы ответственности, у каждого есть владелец

Зачем разделять инфраструктуру на проекты

- Понимать, кто за что платит
- Разграничивать по продуктам, окружениям (production/stage), бизнес-единицам
- Упрощать алерты, отчеты и аудит

[БЕЗ ПРОЕКТОВ — НЕТ ОТВЕТСТВЕННОСТИ,
БЕЗ ОТВЕТСТВЕННОСТИ — НЕТ ПРОЗРАЧНОСТИ]

Как обеспечить прозрачность

Разделять инфраструктуру на отдельные проекты

Наименования должны быть понятными, чтобы определять их предназначение

Общие ресурсы вроде NAT и shared DB выносят в отдельный проект

Reports New

Services report

The report shows the amount of services provided by the project for the calendar month.
Total resources show amounts for non-project services.

2026

All projects

February 2026

Download CSV

KZT 405,397.22

Project	Consumption	Price
▼ 		KZT 100,334.50
▼ 		KZT 7,223.96
▼ 		KZT 127,381.45
▼ 		KZT 170,457.31
Paid by Main balance		KZT 405,397.22

Работа с рисками

Deferred payment New

If there are not enough funds on the balance to pay, resources will continue to be provided on loan without blocking for the period of deferred payment. After that they will be blocked and deleted.

[Learn more.](#)

Available services for deferred payment 2 services 

Deferred payment period  4 days excluding weekends

Service	Grace period	Planned removal 
No active deferred payment		

БЕЗВЫХОДНЫХ СИТУАЦИЙ
НЕ БЫВАЕТ.

ЕСЛИ ВОЗНИКНУТ ТРУДНОСТИ,
МЫ ВСЕГДА НА СВЯЗИ

Инструменты контроля

КОНТРОЛЬ

- История операций
- Текущая стоимость
- Графики расходов

ПРОЗРАЧНОСТЬ

- Проекты
- Отчеты по оказанным услугам

ПРОГНОЗИРУЕМОСТЬ

- Горизонт баланса в днях

РАБОТА С РИСКАМИ

Продуктовые фиши в Servercore

Заморозка VM
в нерабочие часы

Karpenter для оптимизации подов
и нод в Kubernetes

Terraform-провайдеры

Сетевые диски

Тарифные скидки
при заказе ресурсов
от трех месяцев для
выделенных серверов

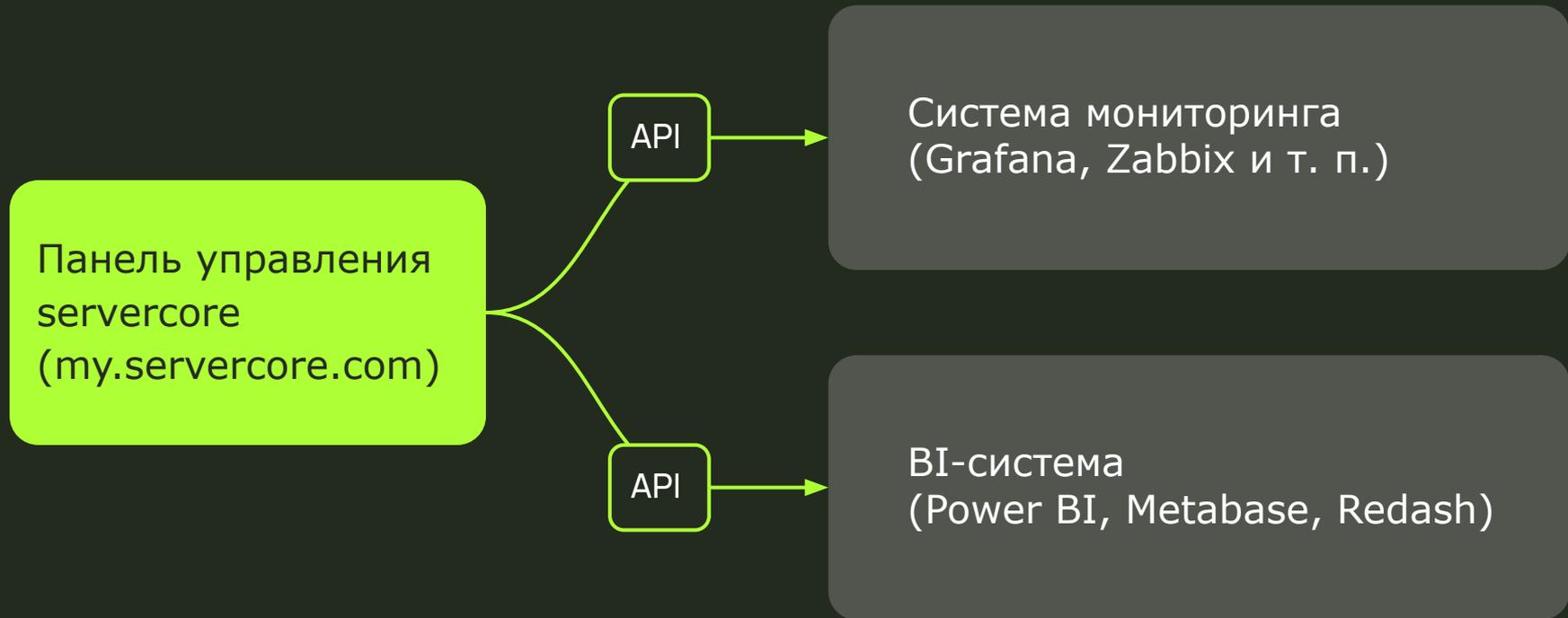
[С ПОМОЩЬЮ ЭТИХ РЕШЕНИЙ МОЖНО ЭФФЕКТИВНЕЕ
УПРАВЛЯТЬ РАСХОДАМИ НА ИНФРАСТРУКТУРУ]

ПЛЮСЫ АВТОМАТИЗАЦИИ

Сценарии расходов на инфраструктуру

[БЕЗ АВТОМАТИЗАЦИИ УПРАВЛЕНИЕ БОЛЬШИМ КОЛИЧЕСТВОМ РЕСУРСОВ ПРЕВРАЩАЕТСЯ В ПРОБЛЕМУ.]

- Длительные ручные проверки
- Высокий риск ошибок
- Запоздалая реакция на перерасход бюджета
- Автоматизация дешевле человеческого ресурса

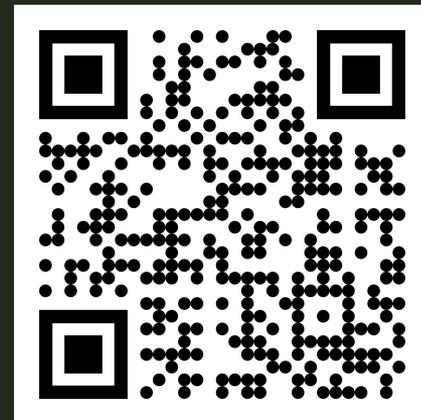


Используйте автоматизацию

С ЕЕ ПОМОЩЬЮ ВЫ СМОЖЕТЕ:

- анализировать данные в реальном времени
- получать быстрые алерты и уведомления
- сокращать время работы команды
- снижать количество ошибок
- минимизировать «сюрпризы» в счетах

API ДОКУМЕНТАЦИИ



Автоматизация — это просто

API → Grafana/Metabase

Slack/Telegram/Email → Правила реакции на сигналы

[БОЛЬШАЯ КОМАНДА НЕ НУЖНА:
ПЕРВАЯ ВЕРСИЯ СОБИРАЕТСЯ БЫСТРО]

API ДОКУМЕНТАЦИИ

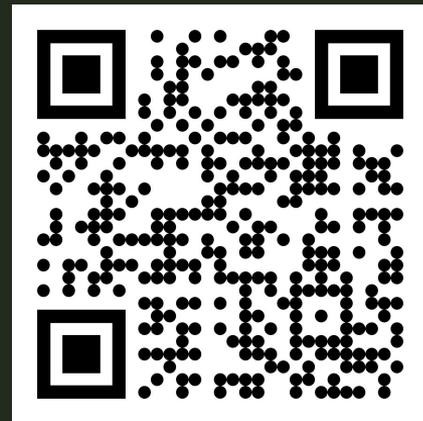


ГРАФИК ПО ДНЯМ МЕСЯЦА С ОДНИМ ПИКОМ ВЫШЕ НА 30% ОТ МЕДИАНЫ



СИГНАЛ

- Скачок дневного расхода на 30% выше медианы
- Всплески по конкретным метрикам

РЕАКЦИЯ

Обнаружили аномалию

Определили продукт и услугу

Сообщили владельцу

Приняли решение: остановить, уменьшить или отложить

Сделали ретроспективу

С ПОМОЩЬЮ ПРОСТЫХ ШАГОВ
МОЖНО ПОВЫСИТЬ КОНТРОЛЬ,
ПРЕДСКАЗУЕМОСТЬ
И ПРОЗРАЧНОСТЬ ЗАТРАТ
НА ИНФРАСТРУКТУРУ

servercore.com

blog.servercore.com

Спасибо
за внимание!

