

Koszt i złożoność operacyjna

Koszt jednostkowy, pojemność i zasoby przerywalne

mgr inż. Jakub Woźniak

Zarządzanie Systemami Rozproszonymi · Semestr zimowy 2026/27

Składnik / miesiąc	Koszt	Udział
Baza danych	1200 EUR	40%
Workery	900 EUR	30%
Telemetria	600 EUR	20%
Sieć i pozostałe	300 EUR	10%
Razem	3000 EUR	100%

Wszystkie kwoty i obciążenia w tym wykładzie: model dydaktyczny, nie cennik dostawcy.

100 000 poprawnie zrealizowanych zamówień

$3000 \text{ EUR} / 100\,000 = \mathbf{0,03 \text{ EUR za zamówienie.}}$

- W kolejnym miesiącu: $3600 \text{ EUR} / 150\,000 = \mathbf{0,024 \text{ EUR.}}$
- Koszt całkowity wzrósł o 20%, jednostkowy spadł o 20%.
- Ten sam zakres kosztów i definicja poprawnej realizacji w obu okresach.

Koszt platformy: 400 EUR	Zespół A	Zespół B
Równy podział	200 EUR	200 EUR
Udział w zużyciu: 75% / 25%	300 EUR	100 EUR

- Wybrana metoda wpływa na decyzje zespołów.
- Tagi wskazują właściciela; nie rozwiązują automatycznie podziału kosztu wspólnego.
- Oddzielić koszty przypisane bezpośrednio od rozliczonych umownie.

Pomiary pod obciążeniem

Szczyt: 100 zadań/s. Jeden worker: 40 zadań/s przy akceptowanym opóźnieniu.
Założenie: brak wąskiego gardła w bazie.

Liczba workerów	Normalnie	Po utracie jednego
4	160 zadań/s	120 zadań/s
3	120 zadań/s	80 zadań/s

- Trzy repliki wystarczają normalnie; nie wystarczają przy założonej awarii.

Model stałych przepływów

Zaległość $B = 6000$ zadań; napływ $\lambda = 100/\text{s}$; obsługa $\mu = 120/\text{s}$. Czas $\approx B / (\mu - \lambda) = \mathbf{300 \text{ s}}$.

- Warunek: $\mu > \lambda$; przy $\mu \leq \lambda$ kolejka nie zostanie opróżniona.
- Chwilowy koszt dodatkowych workerów może ograniczyć czas pogorszenia.
- Model pomija zmienność zadań i ograniczenia zależności.

Obliczenie dydaktyczne bilansu kolejki.



- Odebranie instancji przed ACK → możliwe ponowne dostarczenie.
- Potrzebne idempotencja i odzyskiwalny postęp długiej pracy.
- Dostępność zastępczej instancji Spot nie jest gwarantowana.

Microsoft — Build workloads with Spot VMs

Wariant	Pełna pojemność	Po utracie wszystkich Spot
2 regular + 2 Spot	160/s	80/s
3 regular + 1 Spot	160/s	120/s

- Przy napływie 100/s drugi wariant zachowuje nadwyżkę przepustowości.
- Rabat Spot dotyczy jego części kosztu, nie całego systemu.
- Kryterium: przetestowany czas reakcji i skutek dla użytkownika.

Model

20 GB logów/dzień × 30 dni × 1 EUR/GB ingest = **600 EUR/miesiąc**.

- Ograniczenie ingest do 10 GB/dzień → oszczędność 300 EUR.
- Krótsza retencja nie zmniejsza automatycznie opłaty za ingest.
- Redukować zbędne dane; zachować materiał wymagany do diagnozy i kontroli.

Model obejmuje wyłącznie ingest, bez kosztu przechowywania i zapytań.

Miesiąc — przykład	Własny komponent	Usługa zarządzana
Infrastruktura	300 EUR	700 EUR
Praca operacyjna	$12 \text{ h} \times 50 \text{ EUR} = 600 \text{ EUR}$	$3 \text{ h} \times 50 \text{ EUR} = 150 \text{ EUR}$
Razem w modelu	900 EUR	850 EUR

- Założenia o nakładzie sprawdzić na rzeczywistych zadaniach.
- Migracja, dyżury, ryzyko i ograniczenia funkcji mogą zmienić wynik.

Założenia

Utrzymanie platformy: 80 h/miesiąc. Oszczędność: 10 h/miesiąc na zespół. Próg równowagi: **8 zespołów**.

- Przy 5 zespołach: 50 h oszczędności wobec 80 h utrzymania.
- Model nie obejmuje kosztu początkowego i korzyści z ograniczenia błędów.
- Pomiar: czas dostarczenia zmiany, zgłoszenia, niezawodność i użycie platformy.

Dostępne działania

Workery tańsze o 30%; telemetria tańsza o 50%. Pozostałe koszty bez zmian.
Bazowy rachunek: 3000 EUR.

- Ile oszczędzimy łącznie? Czy osiągniemy cel 30%?
- Jakich testów wymagają obie zmiany?
- Czy redukcja liczby poprawnych realizacji może „poprawić” rachunek?

- FinOps Foundation — Unit Economics
- FinOps Foundation — Allocation
- Microsoft — Build workloads with Spot VMs
- Google SRE — Managing load

Dokumentacja sprawdzona: 26.09.2026. Dane przypadków: przykłady dydaktyczne.