Wydajne i oszczędne architektury

Domeny 3 i 4: wybór usług, modele zakupu i narzędzia kosztowe.

- Wydajne architektury: obliczenia, magazyn, bazy, CloudFront i strumienie danych
- Optymalizacja kosztów: modele zakupu, Savings Plans, cykl życia S3 i narzędzia kosztowe

Wydajne architektury: obliczenia, magazyn, bazy, CloudFront i strumienie danych

Poziom: średnio

Dotyczy: SAA-C03 domena 3, zadania 3.1 do 3.5 (magazyn, obliczenia, bazy, sieć, pozyskiwanie i przekształcanie danych), Amazon EC2, AWS Lambda, Amazon ECS, AWS Fargate, Amazon EKS, Amazon EBS, Amazon EFS, Amazon FSx, Amazon S3, Amazon DynamoDB, Amazon ElastiCache, Amazon CloudFront, AWS Global Accelerator, Amazon Kinesis, AWS Glue, Amazon Athena, dni 18 do 21 planu

Czas: 4 dni po 1,5 do 2 godzin

Domena wydajnych architektur to 24 procent egzaminu i w praktyce pytanie o wybór: które obliczenia, który magazyn, która baza, jak przybliżyć treść do użytkownika i jak przyjąć strumień danych. Ta strona to przede wszystkim tabele wyboru ze słowami kluczami z pytań, a pod nimi laboratorium i krok projektu sklepu.

ZAŁOŻENIA SCENARIUSZA

ćwiczenia robimy w VPC laboratorium 10.60.0.0/16 z bramą NAT włączoną na czas ćwiczeń z kontenerami. Klaster EKS nie jest częścią laboratorium, bo płaszczyzna sterowania kosztuje od godziny, a jego zakładanie i porównanie z AKS opisuje strona Nomad, Podman z Quadlet i zarządzany Kubernetes: AKS i EKS.

Skróty użyte na tej stronie (35)

Skrót	Co znaczy
AD	Active Directory: katalog Microsoftu z kontami użytkowników, komputerów i grup w firmie
AKS	Azure Kubernetes Service: Kubernetes w chmurze Microsoftu, w którym płaszczyznę sterowania utrzymuje dostawca

Skrót	Co znaczy
ALB	Application Load Balancer: równoważnik AWS warstwy 7, kieruje żądania HTTP i HTTPS według ścieżki i nazwy hosta
API	Application Programming Interface: sposób, w jaki programy rozmawiają ze sobą
AWS	Amazon Web Services: chmura Amazonu
CLI	Command Line Interface: praca przez wpisywanie komend w terminalu
CSV	w Windowsie: Cluster Shared Volume, dysk widoczny naraz dla wszystkich węzłów klastra; w danych: Comma-Separated Values, plik z wartościami oddzielonymi przecinkami
DAX	DynamoDB Accelerator: pamięć podręczna przed DynamoDB, skraca odczyty do mikrosekund
DMS	Database Migration Service: usługa AWS do migracji baz danych z replikacją zmian do dnia przełączenia
DNS	Domain Name System: książka telefoniczna internetu, zamienia nazwy na adresy IP
DynamoDB	baza danych typu klucz-wartość w chmurze AWS; Lore trzyma w niej fragmenty, stan, magazyn zmienny i blokady, gdy działa na AWS
EBS	Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny
EC2	Elastic Compute Cloud: usługa AWS wynajmująca gotowe maszyny wirtualne na godziny
ECS	Elastic Container Service: usługa AWS do uruchamiania kontenerów, na własnych instancjach albo na Fargate
EFS	Elastic File System: współdzielony system plików NFS w AWS, dostępny z wielu instancji i stref
EKS	Elastic Kubernetes Service: Kubernetes w chmurze Amazon Web Services, w którym płaszczyznę sterowania utrzymuje dostawca
FSx	rodzina zarządzanych systemów plików AWS: FSx for Windows File Server, Lustre, NetApp ONTAP i OpenZFS
GB	Gigabyte: miliard bajtów, jednostka pojemności pamięci i plików, tu najczęściej rozmiar modelu albo pamięci karty graficznej
IOPS	Input/Output Operations Per Second: liczba operacji odczytu i zapisu na sekundę, podstawowa miara szybkości magazynu, którą podaje też Health Service S2D
NAT	Network Address Translation: podmiana adresów IP na granicy sieci
NFS	Network File System: udostępnianie katalogów po sieci w świecie Linuksa
OAC	Origin Access Control: mechanizm CloudFront, który pozwala czytać prywatny kubełek S3 tylko przez dystrybucję
OSS	Open Source Software: oprogramowanie o otwartym kodzie; w AWS oznaczenie silnika Redis OSS w ElastiCache
PDF	Portable Document Format: format dokumentu, który wygląda tak samo na każdym komputerze, np. zakres akredytacji jednostki na stronie PCA
RDS	Remote Desktop Services: zdalne pulpity i aplikacje Windows Server

Skrót	Co znaczy
S3	Simple Storage Service: magazyn plików w chmurze AWS
SAA-C03	kod wersji egzaminu AWS Solutions Architect Associate zdawanej od 30 sierpnia 2022, aktualnej w październiku 2026
SCT	Seria cęgów prądowych (SCT-013), które mierzą prąd przez obejmowanie przewodu, bez rozcinania obwodu
SNS	Simple Notification Service: usługa AWS, która rozsyła jeden komunikat do wielu odbiorców naraz (kolejki, e-mail, SMS)
SQL	Structured Query Language: język zapytań do relacyjnych baz danych
SQS	Simple Queue Service: kolejka komunikatów AWS, która rozdziela nadawcę i odbiorcę, żeby jeden nie czekał na drugiego
TCP	Transmission Control Protocol: protokół sieciowy z potwierdzaniem dostarczenia
TiB	tebibajt: 2 do potęgi 40 bajtów, jednostka, w której liczą Windows i PowerShell (1TB w PowerShellu); dysk z etykietą 2 TB ma około 1,82 TiB
UDP	User Datagram Protocol: szybki protokół sieciowy bez potwierdzeń
VPC	Virtual Private Cloud: prywatna sieć wirtualna klienta w chmurze AWS, w której stawiamy maszyny i zakładamy tablice tras

Fiszka

Komendy wykonujemy w CloudShell z profilem konta laboratorium. `NazwaFunkcji`, `ArnRoli`, `ArnKolejki`, `NazwaKlastra`, `IdPodsieci`, `IdGrupy`, `IdSystemuPlikow`, `NazwaTabeli`, `IdDystrybucji` i podobne zastępujemy własnymi.

□□ TYLKO ODCZYT

Lista typów instancji z procesorem Graviton: aws ec2 describe-instance-types

```
aws ec2 describe-instance-types --filters Name=processor-info.supported-
architecture,Values=arm64 Name=current-generation,Values=true --query
"InstanceTypes[?starts_with(InstanceType,
'm8g')].[InstanceType,VCpuInfo.DefaultVCpus,MemoryInfo.SizeInMiB]" --output table
```

Instancje z procesorami AWS Graviton (litera `g` w nazwie rodziny) są według AWS tańsze w przeliczeniu na wydajność, ale wymagają obrazów i programów dla architektury arm64. Dostępność rodziny `m8g` w regionie sprawdza właśnie to polecenie.

AWS = Amazon Web Services: chmura Amazonu · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Utworzenie grupy rozmieszczenia: aws ec2 create-placement-group

```
aws ec2 create-placement-group --group-name NazwaGrupyRozmieszczenia --strategy cluster
```

`cluster` stawia instancje blisko siebie w jednej strefie (najmniejsze opóźnienie, obliczenia równoległe), `spread` na osobnym sprzęcie (najwyżej siedem instancji na strefę w grupie), `partition` w odseparowanych partycjach (Hadoop, Cassandra, Kafka).

aws = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Utworzenie funkcji Lambda: aws lambda create-function

```
aws lambda create-function --function-name NazwaFunkcji --runtime python3.13 --role ArnRoli --handler app.handler --zip-file fileb://funkcja.zip --memory-size 512 --timeout 30
```

Pamięć ustawiamy od 128 MB do 10 240 MB, a moc procesora rośnie razem z nią. Maksymalny czas działania funkcji to 15 minut, więc dłuższe zadania idą do Fargate, AWS Batch albo Step Functions. Dostępne wersje środowisk wykonawczych sprawdzamy w dokumentacji Lambda.

AWS = Amazon Web Services: chmura Amazonu · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Podpięcie kolejki SQS jako źródła zdarzeń: aws lambda create-event-source-mapping

```
aws lambda create-event-source-mapping --function-name NazwaFunkcji --event-source-arn ArnKolejki --batch-size 10
```

Lambda sama odpytuje kolejkę, wywołuje funkcję z paczką do dziesięciu komunikatów i usuwa je po sukcesie. Czas niewidoczności kolejki ustawiamy na co najmniej sześciokrotność czasu działania funkcji, jak zaleca dokumentacja Lambda.

□□ **SQS** = Simple Queue Service: kolejka komunikatów AWS, która rozdziela nadawcę i odbiorcę, żeby jeden nie czekał na drugiego · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Utworzenie klastra ECS: aws ecs create-cluster

```
aws ecs create-cluster --cluster-name NazwaKlastra
```

□□ **ECS** = Elastic Container Service: usługa AWS do uruchamiania kontenerów, na własnych instancjach albo na Fargate · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Rejestracja definicji zadania Fargate: aws ecs register-task-definition

```
aws ecs register-task-definition --cli-input-json file://zadanie-www.json
```

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ PLIK KONFIGURACYJNY

Definicja zadania z kontenerem nginx: zadanie-www.json

```
{
  "family": "lab-www",
  "networkMode": "awsvpc",
  "requiresCompatibilities": ["FARGATE"],
  "cpu": "256",
  "memory": "512",
  "executionRoleArn": "arn:aws:iam::IdKonta:role/ecsTaskExecutionRole",
  "containerDefinitions": [
    {
      "name": "www",
      "image": "public.ecr.aws/docker/library/nginx:1.28-alpine",
      "essential": true,
      "portMappings": [ { "containerPort": 80, "protocol": "tcp" } ]
    }
  ]
```

```
}
```

Obraz z przypiętą wersją, nigdy `latest`. Rola wykonawcza pozwala Fargate pobrać obraz i pisać logi. Kombinacje `cpu` i `memory` są ustalone (dla 256 jednostek procesora 512 MB do 2 GB), szczegóły w dokumentacji ECS.

GB = Gigabyte: miliard bajtów, jednostka pojemności pamięci i plików, tu najczęściej rozmiar modelu albo pamięci karty graficznej · **ECS** = Elastic Container Service: usługa AWS do uruchamiania kontenerów, na własnych instancjach albo na Fargate

ZMIENIA SYSTEM

Uruchomienie usługi Fargate za równoważnikiem: aws ecs create-service

```
aws ecs create-service --cluster NazwaKlastra --service-name lab-www --task-definition
lab-www --desired-count 3 --launch-type FARGATE --network-configuration
"awsvpcConfiguration={subnets=[IdPodsieciAppA,IdPodsieciAppB,IdPodsieciAppC],securityGro
ups=[IdGrupyAplikacji],assignPublicIp=DISABLED}" --load-balancers
targetGroupArn=ArnGrupyDocelowej,containerName=www,containerPort=80
```

Dla Fargate grupa docelowa musi mieć typ celu `ip`, a nie `instance`, bo każde zadanie ma własny interfejs sieciowy.

aws = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Utworzenie dysku gp3 z dodatkową wydajnością: aws ec2 create-volume

```
aws ec2 create-volume --availability-zone eu-central-1a --size 100 --volume-type gp3 --
iops 6000 --throughput 250
```

gp3 daje bez dopłaty 3000 IOPS i 125 MiB/s niezależnie od rozmiaru, a więcej dokupujemy osobno. Dysk EBS istnieje w jednej strefie i podpina się do instancji w tej samej strefie.

IOPS = Input/Output Operations Per Second: liczba operacji odczytu i zapisu na sekundę, podstawowa miara szybkości magazynu, którą podaje też Health Service S2D · **EBS** = Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Zmiana typu dysku bez odłączania: aws ec2 modify-volume

```
aws ec2 modify-volume --volume-id IdDysku --volume-type gp3
```

Zmiana gp2 na gp3 działa na żywo i zwykle obniża koszt. Następna zmiana tego samego dysku jest możliwa po zakończeniu poprzedniej i odczekaniu czasu podanego w dokumentacji EBS.

□□ **EBS** = Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Utworzenie systemu plików EFS: aws efs create-file-system

```
aws efs create-file-system --performance-mode generalPurpose --throughput-mode elastic
--encrypted --tags Key=Name,Value=lab-efs
```

□□ **EFS** = Elastic File System: współdzielony system plików NFS w AWS, dostępny z wielu instancji i stref · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Punkt montowania EFS w podsieci: aws efs create-mount-target

```
aws efs create-mount-target --file-system-id IdSystemuPlikow --subnet-id IdPodsieciAppA
--security-groups IdGrupyEfs
```

Punkt montowania zakładamy w każdej strefie, z której montujemy. Grupa zabezpieczeń EFS wpuszcza NFS (TCP 2049) z grupy aplikacji.

□□ **EFS** = Elastic File System: współdzielony system plików NFS w AWS, dostępny z wielu instancji i stref · **NFS** = Network File System: udostępnianie katalogów po sieci w świecie Linuksa ·**TCP** = Transmission Control Protocol: protokół sieciowy z potwierdzaniem dostarczenia ·**aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Utworzenie tabeli DynamoDB na żądanie: aws dynamodb create-table

```
aws dynamodb create-table --table-name NazwaTabeli --attribute-definitions AttributeName=klient,AttributeType=S AttributeName=data_zamowienia,AttributeType=S --key-schema AttributeName=klient,KeyType=HASH AttributeName=data_zamowienia,KeyType=RANGE --billing-mode PAY_PER_REQUEST
```

Klucz partycji (`HASH`) rozkłada dane na partycje, klucz sortowania (`RANGE`) układa elementy w partycji. Tryb na żądanie nie wymaga planowania pojemności i jest dobry dla nieprzewidywalnego ruchu.

DynamoDB = baza danych typu klucz-wartość w chmurze AWS; Lore trzyma w niej fragmenty, stan, magazyn zmienny i blokady, gdy działa na AWS · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Automatyczne wygasanie elementów: aws dynamodb update-time-to-live

```
aws dynamodb update-time-to-live --table-name NazwaTabeli --time-to-live-specification Enabled=true,AttributeName=wygasa
```

Elementy z polem `wygasa` (czas uniksowy w sekundach) w przeszłości DynamoDB usuwa sam, bez opłaty za zapis. Dobre dla sesji i koszyków.

DynamoDB = baza danych typu klucz-wartość w chmurze AWS; Lore trzyma w niej fragmenty, stan, magazyn zmienny i blokady, gdy działa na AWS · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Utworzenie bezserwerowej pamięci podręcznej: aws elasticache create-serverless-cache

```
aws elasticache create-serverless-cache --serverless-cache-name lab-sesje --engine valkey --subnet-ids IdPodsieciAppA IdPodsieciAppB --security-group-ids IdGrupyCache
```

ElastiCache Serverless skaluje się sam i jest rozłożony na strefy. Silnik Valkey to otwarta odnoga Redisa, którą AWS oferuje obok Redis OSS i Memcached; dostępne wersje sprawdzamy w dokumentacji.

AWS = Amazon Web Services: chmura Amazonu · **OSS** = Open Source Software: oprogramowanie o otwartym kodzie; w AWS oznaczenie silnika Redis OSS w ElastiCache · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Unieważnienie plików w CloudFront: aws cloudfront create-invalidation

```
aws cloudfront create-invalidation --distribution-id IdDystrybucji --paths "/zdjecia/*"
```

Każe punktom brzegowym pobrać pliki ze źródła na nowo. Przy częstych zmianach lepiej wersjonować nazwy plików (`logo-v2.png`) niż unieważniać.

aws = AWS CLI: narzędzie chmury Amazonu

TYLKO ODCZYT

Włączenie przyspieszonego przesyłu do kubełka: aws s3api put-bucket-accelerate-configuration

```
aws s3api put-bucket-accelerate-configuration --bucket NazwaKubelka --accelerate-configuration Status=Enabled
```

S3 Transfer Acceleration przyjmuje plik w najbliższym punkcie brzegowym i przesyła siecią AWS. Pomaga przy wysyłce z daleka, płatny za gigabajt.

S3 = Simple Storage Service: magazyn plików w chmurze AWS · **AWS** = Amazon Web Services: chmura Amazonu · **aws** = AWS CLI: narzędzie chmury Amazonu

ZMIENIA SYSTEM

Utworzenie strumienia Kinesis na żądanie: aws kinesis create-stream

```
aws kinesis create-stream --stream-name lab-telemetria --stream-mode-details StreamMode=ON_DEMAND
```

Strumień przechowuje rekordy (domyślnie 24 godziny, do 365 dni) i pozwala wielu odbiorcom czytać je niezależnie, w kolejności w obrębie fragmentu.

aws = AWS CLI: narzędzie chmury Amazonu

Z życia wzięte

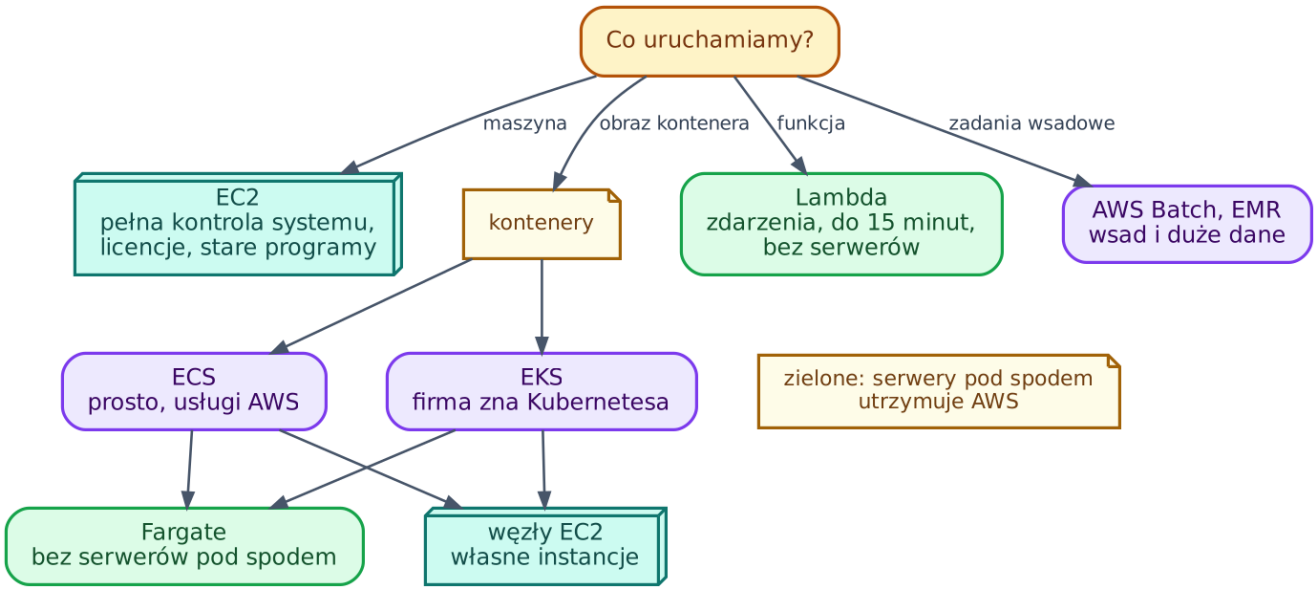
Z ŻYCIA WZIĘTE

Środa, 7:10. Marek przysyła z magazynu zrzut: telemetria wózków ma być widoczna na wykresie w biurze z opóźnieniem najwyżej minuty, a dziś Cassandra na `cass01` do `cass03` dostaje paczki co kwadrans. Tomek w ramach nauki rysuje wersję w AWS: Kinesis Data Streams przyjmuje odczyty, Amazon Data Firehose zapisuje je co minutę do S3 w formacie Parquet, a Athena odpowiada na zapytania Marka zwykłym SQL. Ola pyta, ile to kosztuje przy czterdziestu skanerach, i Tomek dopisuje do dziennika, że od tej pory każda strzałka na rysunku ma mieć liczbę obok.

Jak to działa

Rozwiń: jak to działa pod spodem

Wybór obliczeń



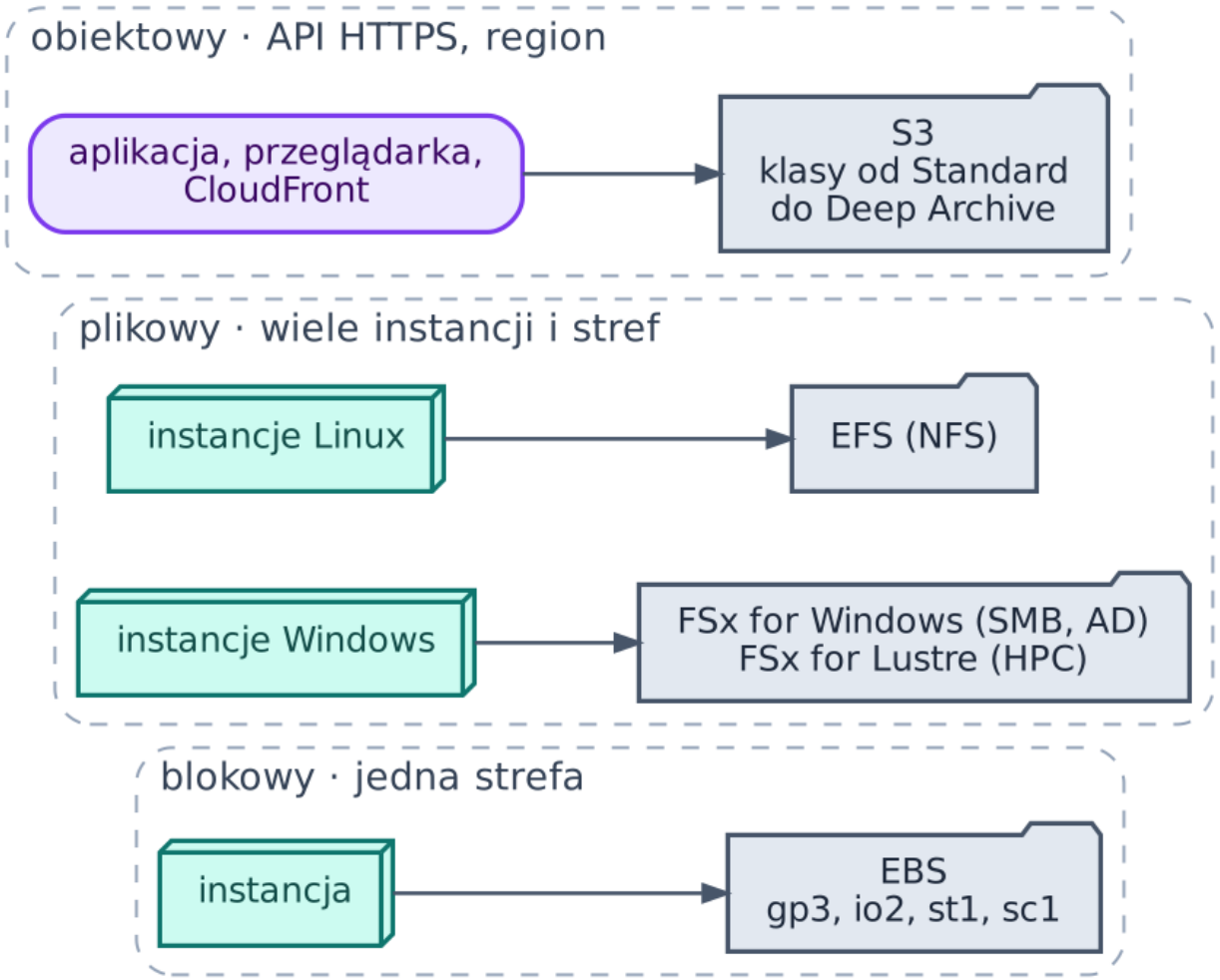
Rysunek 1. Wybór obliczeń w AWS: od pełnej kontroli maszyny w EC2, przez kontenery w ECS i EKS na EC2 albo Fargate, po funkcje Lambda bez serwera; im niżej, tym mniej administracji i mniej kontroli (kliknięcie otwiera pełny rozmiar)

Słowa klucze w pytaniu	Odpowiedź
pełna kontrola systemu, licencja na rdzeń, stary program	EC2
zdarzenia, krótkie zadania do 15 minut, „no servers to manage”	Lambda
kontenery bez zarządzania serwerami	ECS albo EKS na Fargate
kontenery, firma już używa Kubernetesa	EKS
kontenery, prosto, integracja z usługami AWS	ECS
zadania wsadowe z kolejką i dużą skalą	AWS Batch
Hadoop, Spark, przetwarzanie dużych danych	Amazon EMR
obliczenia o małym opóźnieniu między węzłami	EC2 w grupie rozmieszczenia `cluster`
aplikacja WWW bez wiedzy o infrastrukturze	AWS Elastic Beanstalk
AWS w serwerowni klienta	AWS Outposts

Rodzina EC2	Przeznaczenie
M, T (T z kredytami procesora)	ogólne, serwery WWW, małe bazy
C	obliczeniowe: kodowanie wideo, serwery gier, obliczenia naukowe
R, X	pamięciowe: bazy w pamięci, duże pamięci podręczne
I, D	magazynowe: lokalne dyski NVMe o dużej liczbie operacji
P, G, Inf, Trn	akceleratory: GPU, uczenie i wnioskowanie modeli

> **Pojęcie: magazyn instancji (instance store).** Dysk fizycznie w serwerze z instancją, bardzo szybki, ale tracony przy zatrzymaniu albo awarii instancji. Nadaje się na pamięć podręczną i dane tymczasowe, nigdy na jedyną kopię danych.

Wybór magazynu



Rysunek 2. Trzy rodzaje magazynu: blokowy EBS w jednej strefie przy jednej instancji, plikowy EFS i FSx współdzielony przez wiele instancji i stref, obiektowy S3 dostępny przez API z każdego miejsca (kliknięcie otwiera pełny rozmiar)

Słowa klucze	Odpowiedź
dysk dla jednej instancji, baza na EC2	EBS gp3
najwięcej IOPS, baza krytyczna, wiele instancji do jednego dysku	EBS io2 Block Express (Multi-Attach)
duże sekwencyjne odczyty, logi, hurtownia, tanio	EBS st1
rzadko czytane dane na dysku, najtaniej	EBS sc1
system plików Linux współdzielony przez wiele instancji i stref	EFS
udział SMB dla Windows z Active Directory	FSx for Windows File Server

Słowa klucze	Odpowiedź
obliczenia wysokiej wydajności, dane z S3	FSx for Lustre
NetApp, wiele protokołów, migracja z macierzy NetApp	FSx for NetApp ONTAP
obiekty, zdjęcia, kopie, jezioro danych, strona statyczna	S3
serwerownia używa magazynu w AWS jak lokalnego	AWS Storage Gateway (plikowa, wolumenowa, taśmowa)

Liczby EBS według dokumentacji w październiku 2026 roku: gp3 do 64 TiB, 80 000 IOPS i 2000 MiB/s (od września 2025 roku, wcześniej 16 TiB, 16 000 IOPS i 1000 MiB/s), io2 Block Express do 256 000 IOPS i 4000 MiB/s. Część pytań może pamiętać stare limity gp3, więc patrzymy na rząd wielkości: gdy pytanie podaje kilkadziesiąt tysięcy IOPS dla bazy krytycznej, odpowiedzią jest zwykle io2.

S3 obsługuje według dokumentacji co najmniej 3500 zapisów i 5500 odczytów na sekundę na prefiks, więc ruch rozkładamy na wiele prefiksów. Pliki powyżej 100 MB wysyłamy częściami (multipart upload), a CLI robi to sam. S3 Express One Zone to klasa w jednej strefie z opóźnieniem pojedynczych milisekund dla obliczeń, które czytają te same dane wiele razy.

Wybór bazy

Słowa klucze	Odpowiedź
relacyjna, MySQL, PostgreSQL, SQL Server, Oracle, MariaDB, bez zmian w aplikacji	RDS
relacyjna, wysoka wydajność, wiele replik, globalna	Aurora
klucz i wartość, milisekundy przy każdej skali, bez serwerów	DynamoDB
odczyty DynamoDB w mikrosekundach	DynamoDB Accelerator (DAX)
pamięć podręczna przed bazą, sesje, ranking	ElastiCache (Valkey, Redis OSS albo Memcached)
dokumenty JSON, zgodność z MongoDB	Amazon DocumentDB
hurtownia danych, analityka kolumnowa, raporty	Amazon Redshift
graf powiązań	Amazon Neptune
szereg czasowy z czujników	Amazon Timestream
Cassandra bez serwerów	Amazon Keyspaces
migracja bazy z serwerowni	AWS DMS (z SCT przy zmianie silnika)

> **Pojęcie: migracja homogeniczna i heterogeniczna.** Homogeniczna to ten sam silnik po obu stronach (MariaDB do RDS for MariaDB), heterogeniczna to zmiana silnika (Oracle do Aurora PostgreSQL). Druga wymaga przekształcenia schematu narzędziem AWS SCT albo konwersją schematu w DMS.

Sieć blisko użytkownika

Potrzeba	Odpowiedź
treść statyczna i dynamiczna z pamięci brzegowej, HTTP	CloudFront
prywatny kubełek widoczny tylko przez CloudFront	kontrola dostępu do źródła (Origin Access Control)
płatne pliki tylko dla zalogowanych	podpisane adresy URL albo podpisane ciasteczka CloudFront
stałe adresy IP anycast, TCP i UDP, szybkie przełączenie między regionami	AWS Global Accelerator
kod przy punktach brzegowych (przekierowania, nagłówki)	CloudFront Functions albo Lambda@Edge

Pozyskiwanie i przekształcanie danych

Potrzeba	Odpowiedź
strumień w czasie rzeczywistym, wielu odbiorców, powtórne czytanie	Kinesis Data Streams
dostarczenie strumienia do S3, Redshift, OpenSearch bez kodu, z konwersją do Parquet	Amazon Data Firehose (dawniej Kinesis Data Firehose)
Apache Kafka bez utrzymania brokerów	Amazon MSK
zapytania SQL do plików w S3 bez serwera	Amazon Athena
katalog danych i zadania ETL bez serwera	AWS Glue
uprawnienia do jeziora danych na poziomie tabel i kolumn	AWS Lake Formation
wykresy i pulpity dla biznesu	Amazon Quick (dawniej QuickSight)
przesył plików SFTP do S3	AWS Transfer Family

Pułapki egzaminacyjne

- Pliki współdzielone przez instancje w kilku strefach: EFS, nie EBS (dysk jest w jednej strefie, Multi-Attach działa tylko w strefie i tylko z io1 albo io2).
- Aplikacja Windows z udziałami i AD: FSx for Windows File Server, nie EFS (EFS to NFS dla Linuksa).
- „Milisekundy przy dowolnej skali", „bez serwerów", „schemat się zmienia": DynamoDB.
- Ten sam raport czytany tysiące razy obciąża bazę: ElastiCache albo replika odczytu; dla DynamoDB DAX.
- Lambda działa za długo albo potrzebuje więcej niż 15 minut: Fargate albo Step Functions, nie większa pamięć.
- Użytkownicy z całego świata pobierają te same pliki: CloudFront przed S3 z OAC.
- Gra wieloosobowa po UDP z dwóch regionów ze stałymi adresami: Global Accelerator, nie CloudFront.
- Konwersja CSV do Parquet przy pozyskiwaniu: Firehose z konwersją formatu albo Glue.
- Grupa `spread` ma limit instancji na strefę; przy większej liczbie odpowiedzią bywa `partition`.

Częste błędy

Objaw	Przyczyna	Co robimy
Usługa Fargate nie rejestruje zadań w grupie docelowej	grupa docelowa ma typ `instance`	nowa grupa z `--target-type ip`
Zadania Fargate nie startują z błędem pobierania obrazu	brak bramy NAT albo punktów końcowych ECR i S3 w podsieci prywatnej	brama NAT albo punkty końcowe
Instancja nie montuje EFS	grupa EFS nie wpuszcza TCP 2049 albo brak punktu montowania w tej strefie	reguła z grupy aplikacji, punkt montowania w każdej strefie
DynamoDB zwraca błędy przepustowości przy trybie z planowaniem	gorąca partycja albo za mała pojemność	lepszy klucz partycji, tryb na żądanie albo autoskalowanie
Funkcja Lambda kończy się po 3 sekundach	domyślny limit czasu	`--timeout` w `update-function-configuration`
Po zmianie pliku w S3 użytkownicy widzą starą wersję	pamięć CloudFront	wersjonowanie nazw albo unieważnienie

□□ Ćwiczenia

1. Dzień 18. Łączymy kolejkę `lab-zamowienia` ze strony o odpornych usługach z funkcją Lambda, która wypisuje treść komunikatu do dziennika. Jak wygląda kod funkcji i jak sprawdzić, że działa?

Odpowiedź

Plik `app.py` spakowany do `funkcja.zip`:

```
def handler(event, context):
    for rekord in event["Records"]:
        print("zamowienie:", rekord["body"])
    return {"przetworzone": len(event["Records"])}
```

Rola funkcji potrzebuje polityki zarządzanej `AWSLambdaSQSQueueExecutionRole`. Po `create-event-source-mapping` wysyłamy komunikat i czytamy dziennik:

```
aws sqs send-message --queue-url AdresKolejki --message-body "zamowienie-test"
aws logs tail /aws/lambda/NazwaFunkcji --since 5m
```

1. Dzień 19. Uruchamiamy usługę `lab-www` na Fargate za ALB z poprzednich ćwiczeń i porównujemy z grupą Auto Scaling: co zarządzamy sami w każdym wariancie?

Odpowiedź

W grupie Auto Scaling odpowiadamy za obraz systemu, jego aktualizacje, agenta, skrypt startowy i typ instancji. W Fargate odpowiadamy tylko za obraz kontenera, procesor i pamięć zadania oraz sieć, a system pod spodem utrzymuje AWS. W obu wariantach odpowiadamy za aplikację, grupy zabezpieczeń i uprawnienia. Po ćwiczeniu: `aws ecs update-service --cluster NazwaKlastra --service lab-www --desired-count 0`, potem `delete-service` i `delete-cluster`.

1. Dzień 20. Montujemy EFS na dwóch instancjach w różnych strefach i zapisujemy plik z jednej, a czytamy z drugiej. Jakie polecenie montowania na Amazon Linux 2023?

Odpowiedź

```
sudo dnf -y install amazon-efs-utils
sudo mkdir -p /mnt/efs
sudo mount -t efs -o tls IdSystemuPlikow:/ /mnt/efs
echo "z $(hostname)" | sudo tee /mnt/efs/test.txt
```

Na drugiej instancji te same polecenia bez zapisu i `cat /mnt/efs/test.txt`. Opcja `tls` szyfruje ruch NFS pomocnikiem montowania.

1. Pytanie w stylu egzaminu: aplikacja zapisuje miliony odczytów z czujników na sekundę, każdy odczyt ma czytać trzech odbiorców niezależnie, a dane trzeba móc przeczytać ponownie przez siedem dni. Co wybieramy?

Odpowiedź

Kinesis Data Streams z okresem przechowywania wydłużonym do siedmiu dni i trzema aplikacjami odbiorczymi (z rozszerzonym rozsyłaniem, gdy każda ma mieć własną przepustowość). SQS nie pozwala trzem odbiorcom czytać tego samego komunikatu niezależnie ani czytać go ponownie po usunięciu, a SNS nie przechowuje komunikatów.

1. Pytanie w stylu egzaminu: klienci z Azji pobierają z kubełka we Frankfurcie duże katalogi PDF i skarżą się na czas. Co z najmniejszym wysiłkiem?

Odpowiedź

CloudFront z kubełkiem jako źródłem i Origin Access Control, żeby kubełek pozostał prywatny. Replikacja kubełka do Azji też by pomogła, ale wymaga zmian w aplikacji albo DNS i kosztuje więcej. Transfer Acceleration przyspiesza wysyłkę do S3, a nie pobieranie przez wielu odbiorców z pamięci podręcznej.

Krok projektu: sklep B2B

Warstwa aplikacji zostaje na EC2 w grupie Auto Scaling, bo aplikacja Java z `app01` nie jest jeszcze w kontenerze; konteneryzację na ECS z Fargate wpisujemy jako drugi etap. Rodzina instancji `m` z Graviton po teście zgodności, inaczej `m` z procesorem x86. Zdjęcia produktów w S3 za CloudFront z OAC, sesje i koszyki w ElastiCache Serverless. Baza `sklep_b2b` w RDS for MariaDB (strona o danych). Katalog z `mongo01` porównujemy w tabeli: DocumentDB oznacza brak utrzymania, ale zgodność z API MongoDB trzeba sprawdzić dla wersji używanej przez sklep (do potwierdzenia u dostawcy sklepu), a MongoDB na EC2 to pełna zgodność i własna replikacja na trzech instancjach. Migracja bazy przez AWS DMS w trybie homogenicznym z replikacją zmian do dnia przełączenia.

Powiązane: Odporne architektury, część 2: RDS Multi-AZ, Aurora, AWS Backup i strategie DR · Optymalizacja kosztów: modele zakupu, Savings Plans, cykl życia S3 i narzędzia kosztowe · S3: kubełki, obiekty, uprawnienia · Nomad, Podman z Quadlet i zarządzany Kubernetes: AKS i EKS

Źródła: przewodnik SAA-C03, domena 3, zadania 3.1 do 3.5; dokumentacja Amazon EC2 (rodziny instancji, grupy rozmieszczenia, magazyn instancji); dokumentacja Amazon EBS (typy dysków, komunikat z września 2025 o większych gp3); dokumentacja AWS Lambda (limity, źródła zdarzeń SQS); dokumentacja Amazon ECS i AWS Fargate (definicje zadań, typ celu ip); dokumentacja Amazon EFS i Amazon FSx; dokumentacja Amazon S3 (wydajność, multipart, Transfer Acceleration, Express One Zone); blog AWS „Twenty years of Amazon S3” (13 marca 2026); dokumentacja DynamoDB, ElastiCache, CloudFront, Global Accelerator, Kinesis, Amazon Data Firehose, Athena, Glue. Parametry poleceń sprawdzone z modelem usług AWS CLI (botocore 1.43) 19 października 2026.

Optymalizacja kosztów: modele zakupu, Savings Plans, cykl życia S3 i narzędzia kosztowe

Poziom: średnio

Dotyczy: SAA-C03 domena 4, zadania 4.1 do 4.4 (koszty magazynu, obliczeń, baz i sieci), modele zakupu EC2, Savings Plans, Reserved Instances, instancje Spot, klasy i cykl życia S3, AWS DataSync, AWS Storage Gateway, AWS Cost Explorer, AWS Budgets, AWS Cost and Usage Report, AWS Compute Optimizer, dni 22 do 25 planu

Czas: 4 dni po 1,5 godziny

Domena kosztów to 20 procent egzaminu, ale słowa „most cost-effective” pojawiają się także w pytaniach z pozostałych domen. Ta strona opisuje, jak płacimy za obliczenia, magazyn, bazy i sieć, jak przenosimy dane do AWS najtaniej i jakimi narzędziami sprawdzamy, gdzie uciekają pieniądze. Na końcu wyceniamy sklep B2B dla pana Zbigniewa.

ZAŁOŻENIA SCENARIUSZA

narzędzia kosztowe uruchamiamy na koncie zarządzającym organizacji laboratorium, bo ono widzi rachunek wszystkich kont. Cen w złotych i dolarach nie podajemy, bo zależą od regionu i zmieniają się; liczymy je w AWS Pricing Calculator w dniu decyzji. Łącze Zielonej do przesyłu danych to światłowód z adresem 203.0.113.10 ze strony Plan 28 dni i laboratorium AWS w Zielonej.

Skróty użyte na tej stronie (21)

Skrót	Co znaczy
ALB	Application Load Balancer: równoważnik AWS warstwy 7, kieruje żądania HTTP i HTTPS według ścieżki i nazwy hosta
API	Application Programming Interface: sposób, w jaki programy rozmawiają ze sobą

Skrót	Co znaczy
AWS	Amazon Web Services: chmura Amazonu
B2B	business to business: sprzedaż i systemy skierowane do innych firm, nie do klientów indywidualnych
CLI	Command Line Interface: praca przez wpisywanie komend w terminalu
DynamoDB	baza danych typu klucz-wartość w chmurze AWS; Lore trzyma w niej fragmenty, stan, magazyn zmienny i blokady, gdy działa na AWS
EBS	Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny
EC2	Elastic Compute Cloud: usługa AWS wynajmująca gotowe maszyny wirtualne na godziny
ECS	Elastic Container Service: usługa AWS do uruchamiania kontenerów, na własnych instancjach albo na Fargate
GLPI	Gestionnaire Libre de Parc Informatique: otwarty system do zgłoszeń, inwentarza i bazy wiedzy IT; w Zielonej działa w kontenerze na docker01
IPv4	Internet Protocol wersja 4: adresy zapisane czterema liczbami z kropkami, na przykład 10.20.10.77
NAT	Network Address Translation: podmiana adresów IP na granicy sieci
RDS	Remote Desktop Services: zdalne pulpity i aplikacje Windows Server
S3	Simple Storage Service: magazyn plików w chmurze AWS
SAA-C03	kod wersji egzaminu AWS Solutions Architect Associate zdawanej od 30 sierpnia 2022, aktualnej w październiku 2026
SNS	Simple Notification Service: usługa AWS, która rozsyła jeden komunikat do wielu odbiorców naraz (kolejki, e-mail, SMS)
SQL	Structured Query Language: język zapytań do relacyjnych baz danych
SQS	Simple Queue Service: kolejka komunikatów AWS, która rozdziela nadawcę i odbiorcę, żeby jeden nie czekał na drugiego
VPC	Virtual Private Cloud: prywatna sieć wirtualna klienta w chmurze AWS, w której stawiamy maszyny i zakładamy tablice tras
VPN	Virtual Private Network: szyfrowany tunel do sieci firmowej
WAF	Web Application Firewall: zapora aplikacji WWW, która filtruje żądania HTTP, na przykład wstrzykiwanie SQL i zalew żądań

Fiszka

Komendy `ce`, `budgets` i `compute-optimizer` wykonujemy na koncie zarządzającym, pozostałe na koncie laboratorium. Cost Explorer trzeba raz włączyć w konsoli, a dane pojawiają się po około

dobie. Daty, nazwy kubełków i identyfikatory zastępujemy własnymi.

□□ TYLKO ODCZYT

Koszt w tym miesiącu według usług: aws ce get-cost-and-usage

```
aws ce get-cost-and-usage --time-period Start=2026-10-01,End=2026-10-19 --granularity
MONTHLY --metrics UnblendedCost --group-by Type=DIMENSION,Key=SERVICE --query
"ResultsByTime[0].Groups[].[Keys[0],Metrics.UnblendedCost.Amount]" --output table
```

Data `End` nie wchodzi do zakresu. Każde wywołanie API Cost Explorer jest płatne (niewielka opłata za żądanie), więc nie wstawiamy go do pętli co minutę.

□□ **API** = Application Programming Interface: sposób, w jaki programy rozmawiają ze sobą · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Koszt według tagu projektu: aws ce get-cost-and-usage --group-by Type=TAG

```
aws ce get-cost-and-usage --time-period Start=2026-10-01,End=2026-10-19 --granularity
MONTHLY --metrics UnblendedCost --group-by Type=TAG,Key=projekt
```

Działa tylko dla tagów włączonych jako tagi alokacji kosztów, i to od chwili włączenia, a nie wstecz.

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Włączenie tagu alokacji kosztów: aws ce update-cost-allocation-tags-status

```
aws ce update-cost-allocation-tags-status --cost-allocation-tags-status
TagKey=projekt,Status=Active
```

Tag musi być wcześniej użyty na jakimś zasobie, inaczej nie pojawi się na liście.

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Rekomendacja Compute Savings Plans: aws ce get-savings-plans-purchase-recommendation

```
aws ce get-savings-plans-purchase-recommendation --savings-plans-type COMPUTE_SP --term-in-years ONE_YEAR --payment-option NO_UPFRONT --lookback-period-in-days THIRTY_DAYS
```

Cost Explorer liczy, ile godzinowego zobowiązania opłaca się kupić na podstawie ostatnich 30 dni. Na koncie laboratorium, z ruchem tylko podczas nauki, rekomendacja zwykle będzie pusta albo bliska zera, i to też jest wynik.

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Rekomendacje zmiany rozmiaru instancji: aws compute-optimizer get-ec2-instance-recommendations

```
aws compute-optimizer update-enrollment-status --status Active --include-member-accounts
aws compute-optimizer get-ec2-instance-recommendations --query "instanceRecommendations[].[instanceArn,finding,recommendationOptions[0].instanceType]" --output table
```

Compute Optimizer trzeba raz włączyć, a pierwsze rekomendacje pojawiają się po zebraniu danych z kilkunastu godzin pracy instancji. Ocenia też grupy Auto Scaling, dyski EBS, funkcje Lambda i usługi ECS na Fargate.

□□ **EBS** = Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny · **ECS** = Elastic Container Service: usługa AWS do uruchamiania kontenerów, na własnych instancjach albo na Fargate · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Historia cen Spot: aws ec2 describe-spot-price-history

```
aws ec2 describe-spot-price-history --instance-types m7g.large --product-descriptions "Linux/UNIX" --start-time 2026-10-18T00:00:00Z --query "SpotPriceHistory[].[AvailabilityZone,SpotPrice,Timestamp]" --output table
```

Ceny Spot różnią się między strefami i zmieniają się powoli według podaży i popytu.

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Uruchomienie instancji Spot: aws ec2 run-instances --instance-market-options

```
aws ec2 run-instances --image-id resolve:ssm:/aws/service/ami-amazon-linux-latest/al2023-ami-kernel-default-x86_64 --instance-type t3.micro --subnet-id IdPodsieci --instance-market-options MarketType=spot --tag-specifications "ResourceType=instance,Tags=[{Key=projekt,Value=lab-aws}]"
```

AWS może odebrać instancję Spot z dwuminutowym ostrzeżeniem, więc uruchamiamy na niej tylko pracę, którą da się przerwać i wznowić.

□□ **AWS** = Amazon Web Services: chmura Amazonu · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ ZMIENIA SYSTEM

Wykrywanie anomalii kosztów: aws ce create-anomaly-monitor

```
aws ce create-anomaly-monitor --anomaly-monitor '{"MonitorName":"uslugi-lab","MonitorType":"DIMENSIONAL","MonitorDimension":"SERVICE"}'
```

Monitor uczy się zwykłego poziomu kosztów każdej usługi i zgłasza skoki. Powiadomienia ustawiamy osobną subskrypcją (`create-anomaly-subscription`) albo w konsoli.

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Reguła cyklu życia dla logów sklepu: aws s3api put-bucket-lifecycle-configuration

```
aws s3api put-bucket-lifecycle-configuration --bucket NazwaKubelka --lifecycle-configuration file://cykl-logi.json
```

□□ **aws** = AWS CLI: narzędzie chmury Amazonu

□□ PLIK KONFIGURACYJNY

Plik reguły cyklu życia: cykl-logi.json

```
{
  "Rules": [
    {
      "ID": "logi-sklepu",
      "Status": "Enabled",
      "Filter": { "Prefix": "logi/" },
      "Transitions": [
        { "Days": 30, "StorageClass": "STANDARD_IA" },
        { "Days": 90, "StorageClass": "GLACIER_IR" },
        { "Days": 180, "StorageClass": "DEEP_ARCHIVE" }
      ],
      "Expiration": { "Days": 730 },
      "NoncurrentVersionExpiration": { "NoncurrentDays": 30 },
      "AbortIncompleteMultipartUpload": { "DaysAfterInitiation": 7 }
    }
  ]
}
```

Logi przechodzą do coraz tańszych klas i znikają po dwóch latach. Stare wersje (przy włączonym wersjonowaniu) usuwamy po 30 dniach, a niedokończone wysyłki częściowe, które zajmują miejsce niewidocznie, po tygodniu. Regułę dla archiwum z Object Lock opisuje strona Zimne archiwum w chmurze: Azure Blob Storage i Amazon S3 Glacier.

□□ **S3** = Simple Storage Service: magazyn plików w chmurze AWS

□□ TYLKO ODCZYT

Opłata za pobieranie po stronie pobierającego: aws s3api put-bucket-request-payment

```
aws s3api put-bucket-request-payment --bucket NazwaKubelka --request-payment-
configuration Payer=Requester
```

Requester Pays przerzuca opłatę za żądania i transfer na konto, które pobiera dane. Pobierający musi być uwierzytelniony w AWS i dodać do żądania potwierdzenie, a anonimowy dostęp przestaje działać.

□□ **AWS** = Amazon Web Services: chmura Amazonu · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Znalezienie nieprzypisanych adresów publicznych: aws ec2 describe-addresses

```
aws ec2 describe-addresses --query
"Addresses[?AssociationId==null].[PublicIp,AllocationId]" --output table
```

Każdy publiczny adres IPv4 jest płatny za godzinę, przypisany czy nie, więc nieużywane zwalniamy poleceniem `aws ec2 release-address --allocation-id IdAdresu`.

□□ **IPv4** = Internet Protocol wersja 4: adresy zapisane czterema liczbami z kropkami, na przykład 10.20.10.77 · **aws** = AWS CLI: narzędzie chmury Amazonu

□□ TYLKO ODCZYT

Znalezienie niepodpiętych dysków EBS: aws ec2 describe-volumes --filters

```
aws ec2 describe-volumes --filters Name=status,Values=available --query
"Volumes[].[VolumeId,Size,VolumeType,CreateTime]" --output table
```

Dysk w stanie `available` nie jest podpięty do żadnej instancji, a płacimy za niego jak za używany.

□□ **EBS** = Elastic Block Store: dyski blokowe maszyn wirtualnych EC2 w chmurze AWS, rozliczane osobno od maszyny · **aws** = AWS CLI: narzędzie chmury Amazonu

Z życia wzięte

□□ Z ŻYCIA WZIĘTE

Poniedziałek, 9:30. Pan Zbigniew dostaje od Tomka pierwszą wycenę sklepu w AWS i pyta, dlaczego pozycja „transfer danych” jest prawie tak duża jak serwery. Tomek otwiera kalkulator i pokazuje, że policzył zdjęcia produktów wysyłane klientom wprost z kubełka, bez CloudFront, i bramę NAT, przez którą instancje pobierają te same zdjęcia z S3. Po dodaniu CloudFront i punktu końcowego bramy dla S3 pozycja spada wyraźnie. Pan Zbigniew mówi, że od teraz chce widzieć w każdej wycenie osobny wiersz na transfer.

Jak to działa

Rozwiń: jak to działa pod spodem

Modele zakupu obliczeń



Rysunek 1. Modele zakupu EC2 od najdroższego i najbardziej elastycznego do najtańszego: na żądanie, Savings Plans i Reserved Instances z zobowiązaniem na rok albo trzy lata, Spot z ryzykiem przerwania, oraz hosty dedykowane dla licencji (kliknięcie otwiera pełny rozmiar)

Model	Rabat (według AWS, „do”)	Zobowiązanie	Kiedy na egzaminie
na żądanie (On-Demand)	brak	brak	krótkie, nieprzewidywalne obciążenia, testy
Compute Savings Plans	do 66 procent	kwota na godzinę przez rok albo trzy lata	stałe obciążenie, które może zmieniać rodzinę, region, przejść na Fargate albo Lambdę
EC2 Instance Savings Plans	do 72 procent	kwota na godzinę, jedna rodzina w jednym regionie	stałe obciążenie, rodzina i region pewne
Reserved Instances (standardowe i wymienialne)	podobne do Savings Plans	konkretny typ na rok albo trzy lata	starsze pytania; strefowe RI dają też rezerwację pojemności
Spot	do 90 procent	brak, AWS może przerwać z dwuminutowym ostrzeżeniem	przetwarzanie wsadowe, CI, zadania odporne na przerwanie

Model	Rabat (według AWS, „do”)	Zobowiązanie	Kiedy na egzaminie
rezerwacja pojemności na żądanie	brak sama w sobie	płacimy od chwili utworzenia	gwarancja dostępności pojemności w strefie na konkretny termin
host dedykowany (Dedicated Host)	brak	opcjonalnie rezerwacja	licencje na gniazdo albo rdzeń fizyczny (BYOL), zgodność
instancja dedykowana	brak	brak	sprzęt niewspółdzielony z innymi klientami, bez widoczności gniazd

Do tych modeli doszły w grudniu 2025 roku Database Savings Plans: do 35 procent rabatu na RDS, Aurorę, DynamoDB, ElastiCache, DocumentDB i inne bazy, z możliwością zmiany silnika i regionu. Przewodnik SAA-C03 ich nie zna, więc w pytaniach o bazy spodziewamy się raczej zarezerwowanych instancji baz (Reserved DB Instances) albo trybu na żądanie i bezserwerowego.

Mieszanka, która często jest najlepszą odpowiedzią: podstawowe obciążenie pokryte Savings Plans, szczyty na żądanie, a zadania wsadowe na Spot. Grupa Auto Scaling umie łączyć te modele polityką mieszanych instancji.

Koszty magazynu i przesyłu danych

Klasa S3	Minimalny okres rozliczany	Odczyt	Kiedy
Standard	brak	od ręki, bez opłaty za odczyt danych	dane używane często
Intelligent-Tiering	brak	od ręki, mała opłata za monitorowanie obiektu	wzorzec dostępu nieznany albo zmienny
Standard-IA	30 dni	od ręki, opłata za gigabajt odczytu	rzadko czytane, ale potrzebne od razu
One Zone-IA	30 dni	od ręki, jedna strefa	dane łatwe do odtworzenia, kopie drugorzędne
Glacier Instant Retrieval	90 dni	milisekundy	archiwum czytane raz na kwartał
Glacier Flexible Retrieval	90 dni	minuty do godzin, po zleceniu	archiwum bez pilnych odczytów
Glacier Deep Archive	180 dni	do 12 godzin w trybie Standard, do 48 w Bulk	wymogi przechowywania na lata

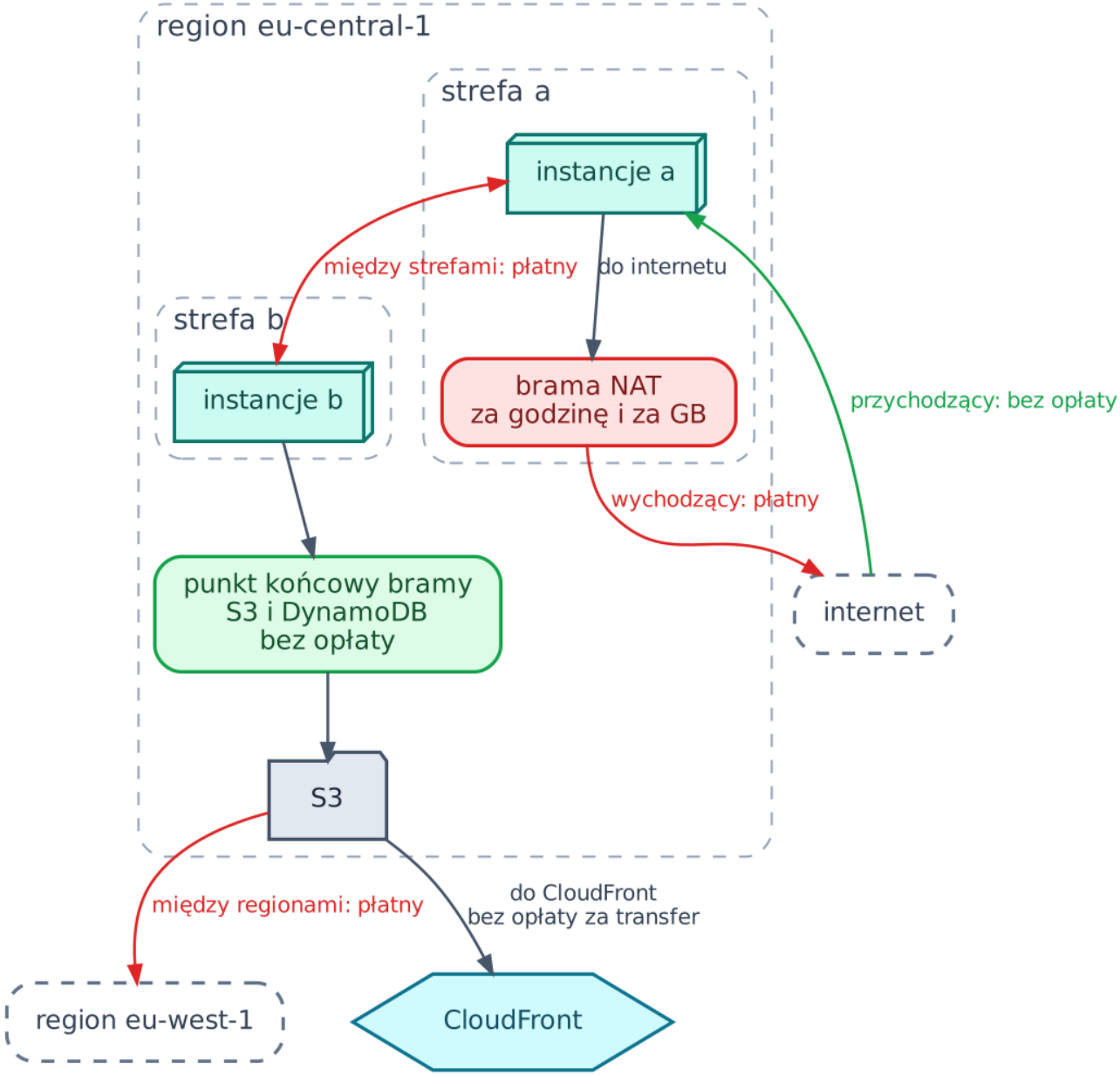
Czasy odtworzenia z Glacier i narzut na małe obiekty opisuje szczegółowo strona Zimne archiwum w chmurze: Azure Blob Storage i Amazon S3 Glacier.

Przesył danych do AWS	Kiedy
AWS DataSync	online, z serwera NFS lub SMB albo innej chmury do S3, EFS albo FSx, z harmonogramem i weryfikacją
AWS Storage Gateway	stały, hybrydowy dostęp z serwerowni do magazynu w AWS (udziały plikowe, wolumeny, wirtualne taśmy)
AWS Transfer Family	partnerzy wysyłają pliki przez SFTP, FTPS albo AS2
S3 Transfer Acceleration	wysyłka do S3 z dalekich miejsc przez sieć AWS
AWS Direct Connect	stałe, duże wolumeny i przewidywalny koszt przesyłu
AWS Data Transfer Terminal	offline: przywozimy własne nośniki do punktu AWS
rodzina AWS Snow (Snowball Edge)	offline: urządzenie wysyłane pocztą; według dokumentacji AWS niedostępne już dla nowych klientów

> □□ Przewodnik SAA-C03 z 2022 roku i większość pytań próbnych traktuje Snowball jako odpowiedź na „setki terabajtów przez wolne łącze”. Dokumentacja AWS mówi dziś, że urządzenia Snow nie są oferowane nowym klientom, a jako zastępstwo wskazuje DataSync, Data Transfer Terminal i rozwiązania partnerów. Na egzaminie wybieramy Snowball, jeśli jest jedyną odpowiedzią offline; w projekcie dla Zielonej liczymy DataSync przez łącze.

Rachunek przesyłu: ile dni zajmie wysyłka przez łącze. 10 TB przez łącze 1 Gb/s wykorzystane w 80 procentach to około 10 × 8000 Gb / 0,8 Gb/s, czyli 100 000 sekund, nieco ponad doba. 500 TB przez to samo łącze to już ponad osiem tygodni, i wtedy odpowiedzią jest przesył offline albo Direct Connect.

Koszty sieci



Rysunek 2. Gdzie płacimy za ruch: wychodzący do internetu, między strefami i regionami, przez bramę NAT za każdy gigabajt; bez opłaty ruch przychodzący i punkt końcowy bramy do S3 i DynamoDB (kliknięcie otwiera pełny rozmiar)

Ruch	Koszt	Jak obniżyć
przychodzący z internetu	bez opłaty	
wychodzący do internetu	płatny za gigabajt	CloudFront, kompresja, pamięć podręczna

Ruch	Koszt	Jak obniżyć
między strefami w regionie	płatny w obie strony	trzymamy rozmowne usługi w tej samej strefie tam, gdzie odporność na to pozwala
między regionami	płatny	replikujemy tylko to, czego wymaga DR
przez bramę NAT	opłata za godzinę i za każdy gigabajt	punkt końcowy bramy dla S3 i DynamoDB, punkty interfejsu dla innych usług przy dużym ruchu
punkt końcowy bramy (S3, DynamoDB)	bez opłaty	
punkt końcowy interfejsu	za godzinę i gigabajt	opłaca się zamiast NAT przy dużym ruchu do jednej usługi
VPC peering i Transit Gateway	peering bez opłaty za połączenie, Transit Gateway za załącznik i gigabajt	peering przy kilku VPC, Transit Gateway przy wielu

Brama NAT w każdej strefie to wyższy koszt stały, ale bez opłat za ruch między strefami i bez pojedynczego punktu awarii. Jedna brama NAT dla środowiska testowego to świadoma oszczędność, którą egzamin akceptuje, gdy pytanie mówi o środowisku nieprodukcyjnym.

Narzędzia kosztowe

Narzędzie	Co daje
AWS Pricing Calculator	wycena przed wdrożeniem
AWS Cost Explorer	wykresy i rozbicie kosztów, prognoza, rekomendacje Savings Plans i RI
AWS Budgets	progi kwot albo użycia z powiadomieniami, akcje budżetu (na przykład podpięcie polityki blokującej)
AWS Cost and Usage Report (w Data Exports)	najbardziej szczegółowe dane rozliczeniowe do S3, analiza w Athena
AWS Cost Anomaly Detection	wykrywanie nietypowych skoków kosztów
AWS Compute Optimizer	rekomendacje rozmiaru EC2, Auto Scaling, EBS, Lambda, Fargate
AWS Trusted Advisor	kontrole kosztów, bezpieczeństwa i limitów; pełna lista przy wyższych planach wsparcia
tagi alokacji kosztów	podział rachunku na projekty i zespoły

Narzędzie	Co daje
AWS Organizations (rozliczenia skonsolidowane)	jeden rachunek, wspólne rabaty ilościowe i współdzielenie Savings Plans między kontami

Pułapki egzaminacyjne

- Stałe obciążenie przez trzy lata, rodzina może się zmienić: Compute Savings Plans. Konkretna rodzina i region pewne: EC2 Instance Savings Plans albo standardowe RI.
- Przetwarzanie nocne, które można przerwać i powtórzyć: Spot. Baza produkcyjna nigdy na Spot.
- Licencja Windows Server albo SQL Server na rdzeń fizyczny: host dedykowany.
- Dostęp do danych nieznany: S3 Intelligent-Tiering, a nie ręczne reguły zgadujące wzorzec.
- Obiekty kasowane po kilku dniach: nie przenosimy ich do Standard-IA ani Glacier, bo minimalny okres i tak naliczy 30 albo 90 dni.
- Dużo ruchu z instancji prywatnych do S3: punkt końcowy bramy zamiast NAT.
- Środowisko testowe w godzinach pracy: harmonogram wyłączania (EventBridge Scheduler albo zaplanowane działania Auto Scaling), a nie Savings Plans.
- Pliki dla partnerów, którzy mają płacić za pobranie: Requester Pays.
- Najtańsza baza dla rzadkiego i zmiennego ruchu: Aurora Serverless v2 albo DynamoDB na żądanie, a nie duża instancja z rezerwacją.

Częste błędy

Objaw	Przyczyna	Co robimy
Grupowanie po tagu zwraca pusty klucz dla większości kosztów	tag nie był włączony jako alokacji kosztów albo zasoby go nie mają	`update-cost-allocation-tags-status`, tagowanie, dane tylko od chwili włączenia
Rachunek za S3 Glacier wyższy niż zakładany	usunięcie albo zmiana klasy przed minimalnym okresem, dużo małych obiektów	reguły cyklu życia z progiem rozmiaru, paczkowanie małych plików
Rachunek za bramę NAT rośnie z ruchem	instancje pobierają dane z S3 przez NAT	punkt końcowy bramy S3
Savings Plans nie obniżył rachunku za instancje dedykowane	opłata regionalna za instancje dedykowane nie jest objęta rabatem	uwzględniamy ją osobno w wycenie
Compute Optimizer nie pokazuje rekomendacji	usługa niewłączona albo za mało danych	`update-enrollment-status`, odczekanie

Objaw	Przyczyna	Co robimy
Spot przerywa zadania w kółko	jeden typ w jednej strefie	wiele typów i stref, strategia alokacji zoptymalizowana pod pojemność

□□ Ćwiczenia

1. Dzień 22. Sklep ma w projekcie stałe trzy instancje przez cały rok i do sześciu dodatkowych przez osiem tygodni sezonu. Jak kupujemy obliczenia?

Odpowiedź

Trzy stałe instancje pokrywamy Compute Savings Plans na rok (kwota na godzinę równa ich kosztowi), bo w drugim etapie aplikacja może przejść na Fargate, który ten plan też obejmuje. Instancje sezonowe na żądanie, z zaplanowanym zwiększeniem grupy w listopadzie. Spot nie, bo instancje obsługują sesje klientów i zamówienia. Wysokość zobowiązania sprawdzamy rekomendacją Cost Explorer po kilku tygodniach działania.

1. Dzień 23. Logi aplikacji sklepu mają być dostępne od ręki przez 30 dni, potem czytane sporadycznie przez rok, a przechowywane dwa lata ze względu na reklamacje. Jak wygląda reguła?

Odpowiedź

Plik `cykl-logi.json` z Fiszki z jedną zmianą: przejście do `GLACIER_IR` po 30 dniach zamiast do `STANDARD_IA`, bo sporadyczny odczyt z Glacier Instant Retrieval jest od ręki i taniej przechowuje dane. Przejście do `DEEP_ARCHIVE` po 395 dniach (30 plus rok), wygaśnięcie po 730 dniach. Gdy logi to wiele małych plików, łączymy je w paczki przed wysyłką, bo przejścia małych obiektów kosztują więcej niż oszczędzają.

1. Dzień 24. Liczymy w Pricing Calculator dwa warianty dla instancji w trzech podsieciach prywatnych, które pobierają miesięcznie 2 TB z S3: trzy bramy NAT oraz trzy bramy NAT z punktem końcowym bramy S3. Która pozycja się zmienia i dlaczego?

Odpowiedź

Zmienia się opłata za przetworzone dane bram NAT: z punktem końcowym bramy ruch do S3 nie przechodzi przez NAT, więc 2 TB nie są naliczane jako przetworzone przez bramę. Opłata godzinowa trzech bram zostaje, bo bramy są nadal potrzebne do internetu. Punkt końcowy bramy jest bezpłatny. Kwoty odczytujemy z kalkulatora w dniu ćwiczenia.

1. Dzień 25. Pokazujemy koszty laboratorium za ostatnie 25 dni według usług i według tagu `projekt`. Które usługi zwykle są na górze listy w takim laboratorium i dlaczego?

Odpowiedź

Zwykle VPC (brama NAT i publiczne adresy IPv4), EC2 (instancje i dyski), Elastic Load Balancing i RDS, bo naliczają się za każdą godzinę istnienia, nawet bez ruchu. S3, Lambda, SQS i DynamoDB na żądanie są w laboratorium blisko zera. Gdy po tagu wychodzi duży koszt bez tagu, szukamy zasobów utworzonych ręcznie.

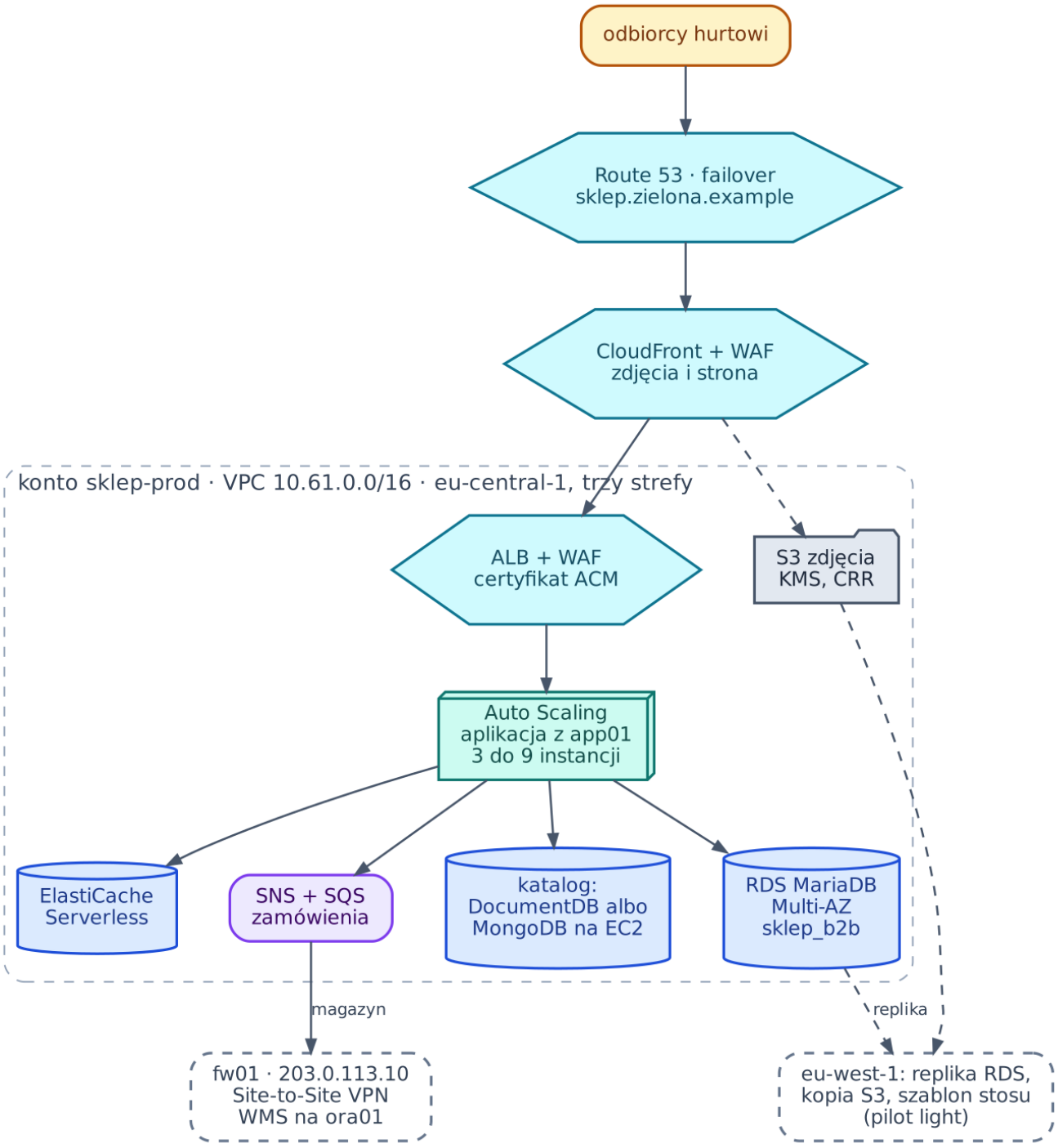
1. Pytanie w stylu egzaminu: firma przetwarza co noc obrazy z kolejki SQS, zadania trwają po kilka minut i mogą być powtórzone, a celem jest najniższy koszt. Co wybieramy?

Odpowiedź

Grupa Auto Scaling na instancjach Spot z kilkoma typami instancji w kilku strefach, skalowana według liczby komunikatów w kolejce, albo AWS Batch na Spot. Komunikat wraca do kolejki po upływie czasu niewidoczności, gdy instancja zostanie odebrana. Instancje na żądanie i Savings Plans są droższe dla pracy, która może być przerwana.

Krok projektu: wycena sklepu B2B dla pana Zbigniewa

Wycenę robimy w AWS Pricing Calculator dla `eu-central-1` i zapisujemy link do niej w zgłoszeniu w GLPI. Rysunek zbiera wszystkie kroki projektu z poprzednich stron w jedną architekturę docelową.



Rysunek 3. Docelowa architektura sklepu B2B w AWS: Route 53 z przełączeniem, CloudFront i WAF, ALB i grupa Auto Scaling w trzech strefach, ElastiCache, RDS MariaDB Multi-AZ, katalog, zamówienia przez SNS i SQS, VPN do fw01 i pilot light w eu-west-1 (kliknięcie otwiera pełny rozmiar)

Wiersze wyceny, każdy z założeniem obok:

Pozycja	Założenie
EC2 aplikacji	3 instancje przez rok z Compute Savings Plans, do 6 na żądanie przez 8 tygodni

Pozycja	Założenie
ALB	jeden, ruch według logów sklepu z listopada
RDS for MariaDB Multi-AZ	jedna instancja, magazyn gp3, kopie 14 dni, Database Savings Plans albo instancja zarezerwowana po pół roku
replika w `eu-west-1`	jedna instancja, transfer replikacji między regionami
ElastiCache Serverless	dane sesji i koszyków
S3 i CloudFront	CloudFront przed całym sklepem (zdjęcia z S3, strony z ALB), transfer wychodzący przez CloudFront
bramy NAT i punkty końcowe	trzy bramy, punkt bramy S3, punkty interfejsu Secrets Manager
Site-to-Site VPN	jedno połączenie z `fw01`, dwa tunele
WAF, Route 53, AWS Backup, CloudWatch	według liczby reguł, stref i danych
przesył danych	jednorazowo DataSync dla zdjęć, DMS dla bazy

Pan Zbigniew dostaje dwie liczby: koszt miesięczny w sezonie i poza nim, oraz porównanie z kosztem utrzymania sklepu w serwerowni (część serwera `app01`, `my01` i `mongo01`, prąd, kopie, nasza praca). Decyzja należy do niego, a Tomek ma po tej wycenie odpowiedź na każde pytanie z domeny kosztów.

Powiązane: Wydajne architektury: obliczenia, magazyn, bazy, CloudFront i strumienie danych · Próbny egzamin SAA-C03 i strategia na dzień egzaminu · Chmura: pojęcia, które trzeba znać zanim cokolwiek uruchomimy

Źródła: przewodnik SAA-C03, domena 4, zadania 4.1 do 4.4; dokumentacja Savings Plans (rodzaje planów i rabaty, w tym Database Savings Plans z grudnia 2025); dokumentacja Amazon EC2 (modele zakupu, Spot, rezerwacje pojemności, hosty dedykowane); dokumentacja Amazon S3 (klasy, minimalne okresy, cykl życia, Requester Pays); dokumentacja AWS Snowball Edge (zmiana dostępności dla nowych klientów); dokumentacja AWS DataSync, Storage Gateway, Transfer Family; dokumentacja Amazon VPC (cennik bram NAT, punktów końcowych, publicznych adresów IPv4); dokumentacja AWS Cost Explorer, Budgets, Cost Anomaly Detection, Compute Optimizer, Data Exports. Parametry poleceń sprawdzone z modelem usług AWS CLI (botocore 1.43) 19 października 2026.