

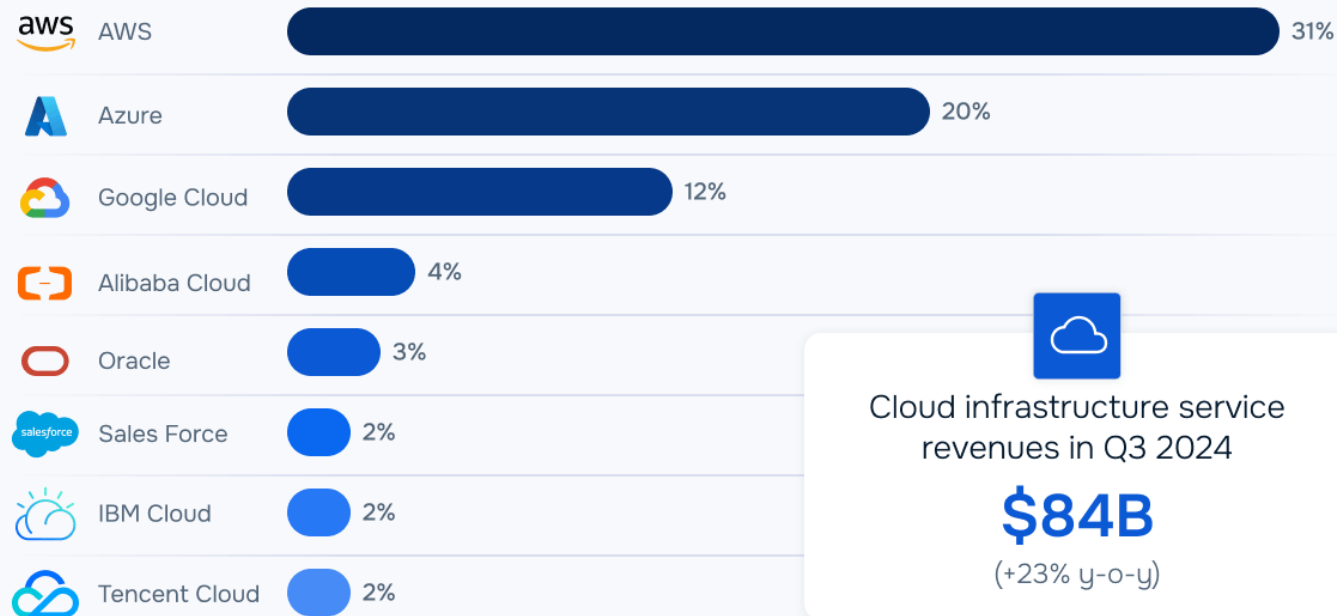


Konrad Sitek

Wprowadzenie do baz NoSQL w AWS

Udział Amazon Web Services w rynku usług chmurowych

Worldwide market share of leading cloud infrastructure service providers in Q3 2024*



Przegląd usług NoSQL w AWS

- DynamoDB
- DocumentDB
- Keyspaces
- Neptune
- ElasticCache
- Timestream
- QLDB



Amazon DynamoDB



Przechowywanie
danych aplikacji w
czasie rzeczywistym



Szybkie
odczyty/zapisy przy
dużym obciążeniu



Koszyki zakupowe,
sesje użytkowników,
profile, IoT

- Bezserwerowa i automatycznie skalująca się
- Opóźnienia w milisekundach
- Klucz-wartość i dokument (JSON)
- Wysoka dostępność (multi-AZ)
- Integracja z AWS Lambda i Streams

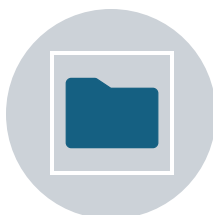
Amazon DocumentDB



PRZECHOWYWANIE
DANYCH W POSTACI
DOKUMENTÓW JSON



ZAMIENNIK MONGODB
W AWS



CMS-Y, KATALOGI
PRODUKTÓW, PROFILE
UŻYTKOWNIKÓW



- Kompatybilna z MongoDB (API + narzędzia)
- Wysoka dostępność i backupy
- Skalowalne odczyty (read replicas)
- W pełni zarządzana
- Stosowana tam, gdzie elastyczny schemat danych

Amazon Keyspaces



Przetwarzanie dużych
ilości danych o
strukturze tabelowej



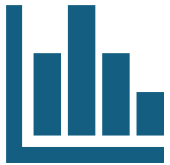
IoT, dane
telemetryczne, logi,
czaty



Wymagające systemy
z ogromnym zapisem

- Kompatybilna z Apache Cassandra (CQL)
- Automatyczna replikacja i skalowanie
- Serwerless i płatność za użycie
- Wysoka dostępność i trwałość danych
- Praca w czasie rzeczywistym

Amazon Neptune



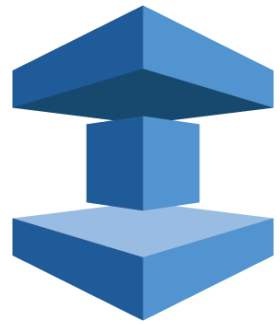
Analiza powiązań i
zależności między danymi



Rekomendacje, wykrywanie
oszustw, sieci
społecznościowe

- Baza grafowa (Gremlin / SPARQL)
- Reprezentacja danych jako wierzchołki i relacje
- Optymalizowana pod szybkie „nawigowanie” w grafie
- Odporna na awarie, skalowalna
- Integracja z ML (Graph ML w SageMaker)

Amazon ElastiCache



SZYBKIE PRZETRZYMYWANIE
TYMCZASOWYCH DANYCH W
RAM



CACHING ZAPYTAŃ, SESJE,
LEADERBOARDY, RATE
LIMITING

- Redis lub Memcached (do wyboru)
- Opóźnienia poniżej 1 ms
- Odciąża inne bazy danych
- Obsługuje replikację i automatyczne przełączenia
- W pełni zarządzany i skalowalny

Amazon Timestream

- Przechowywanie danych czasowych (time series)
- Metryki aplikacji, odczyty czujników, dane IoT



Serverless, bez
potrzeby zarządzania
infrastrukturą



Automatyczne
zarządzanie danymi
„gorącymi” i „zimnymi”



Wbudowane funkcje
agregacji w czasie

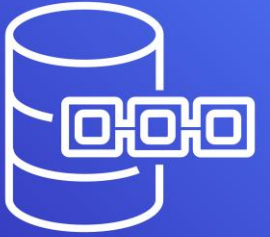


Integracja z Grafana i
Quicksight



Obsługuje miliardy
punktów danych

Amazon QLDB



SYSTEMY WYMAGAJĄCE
HISTORII ZMIAN DANYCH
(LEDGER)



AUDYT, FINANSE, REJESTRY
PUBLICZNE

- Niezmienialny dziennik danych (ledger)
- Kryptograficznie weryfikowalna historia zmian
- W pełni zarządzany, skalowalny
- Szybkie zapytania w modelu dziennika
- Alternatywa dla blockchain w środowiskach zaufanych

Rzeczywiste przykłady firm używających danych usług AWS



DynamoDB – Netflix (zarządzanie stanem odtwarzacza filmów)



Neptune – Thomson Reuters (analiza sieci powiązań prawnych i biznesowych)



DocumentDB – Samsung (zarządzanie katalogami produktów)



ElastiCache – Expedia (cache'owanie wyników wyszukiwania hoteli)



Timestream – Siemens (monitorowanie sprzętu IoT)



QLDB – BMW (rejestracja historii serwisowej pojazdów)

Jak wybrać odpowiednią usługę NoSQL?

Kryteria decyzyjne:

- rodzaj danych (JSON, graf, serie czasowe)
- typ zapytań (proste, złożone, zakresowe)
- wymagania wydajnościowe (czas odpowiedzi, liczba operacji na sekundę)
- potrzeba globalnej dostępności (czy potrzebna replikacja w wielu regionach)

Typowe błędy przy wyborze bazy NoSQL

- stosowanie DynamoDB bez analizy wzorców zapytań (co skutkuje drogimi skanowaniami)
- wybór Neptuna do prostych danych (niepotrzebne komplikowanie)
- użycie DocumentDB zamiast ElastiCache (gdy wymagane ultra-niskie opóźnienia)



Część praktyczna

https://github.com/sit3kk/AWS_DynamoDB



Amazon
DynamoDB