ZAGADNIENIA NA EGZAMIN DYPLOMOWY MAGISTERSKI

Kierunek studiów: **INFORMATYKA**

Studia stacjonarne drugiego stopnia

Specjalność: **Technologie Przetwarzania Danych**

Lp.	Zagadnienie
Big Data i przetwarzanie w chmurze	
1.	Cele i metody fragmentacji, partycjonowania i shardingu danych.
2.	Metody równoważenia obciążenia w chmurach obliczeniowych.
3.	Metody replikacji danych w kontekście poziomów spójności danych.
4.	Algorytmy osiągania kworum i konsensusu na przykładzie protokołu RAFT.
5.	Skracanie czasu odpowiedzi przez zrównoleglanie przetwarzania według prawa Amdahla.
6.	Model przetwarzania Map-Reduce.
Eksploracja danych	
7.	Reguły asocjacyjne: sformułowanie problemu, typy reguł i algorytmy odkrywania reguł asocjacyjnych.
8.	Naiwny klasyfikator Bayesa: ogólna charakterystyka i zasada działania.
9.	Metody indukcji drzew decyzyjnych (miary oceny jakości podziału, metody przycinania drzewa).
10.	Regresja: sformułowanie problemu, rodzaje regresji.
11.	Grupowanie obiektów: sformułowanie problemu, omówienie podstawowych algorytmów grupowania.
12.	Algorytmy redukcji wymiarowości.
Przetwarzanie strumieni danych w systemach Big Data	
13.	Znaczenie punktów kontrolnych w systemach przetwarzania strumieni: zawartość, typy, tworzenie, wykorzystanie.
14.	Systemy wymiany wiadomości: typy, postacie komunikatów, sposób organizacji danych, mechanizmy wpływające na skalowalność i niezawodność.
15.	Typy i rola znaczników czasowych w systemach przetwarzania strumieni danych, nieuporządkowanie strumienie danych, zdarzenia opóźnione, spóźnione, sposób obsługi.
16.	Sterowanie zawartością oraz częstotliwością dostarczania wyników w systemach przetwarzania strumieni danych.
Hurtownie danych i przetwarzanie analityczne	
17.	Architektury systemów hurtowni danych.
18.	Implementacyjne modele hurtowni danych.
19.	Podstawowe schematy hurtowni danych w modelu relacyjnym.

20.	Problematyka projektowania warstwy ETL.
21.	Struktury fizyczne dla hurtowni danych.
22.	Techniki optymalizacji zapytań analitycznych.
23.	Charakterystyka przetwarzania danych w superserwerach - main memory appliance.
24.	Architektury systemów składowania danych masywnych (gigadane, big data).
Technologie dla aplikacji klasy enterprise	
25.	Odwzorowanie obiektowo-relacyjne, JPA, Entity Framework.
26.	Architektura Model-View-Controller w aplikacjach internetowych.
27.	Architektura aplikacji SPA (ang. single-page application).
Administrowanie systemami baz danych	
28.	Charakterystyka punktów kontrolnych.
29.	Porównanie indeksów drzewiastych i indeksów bitmapowych.
Rozproszone bazy danych	
30.	Porównanie fragmentacji poziomej i pionowej.
31.	Metody wykrywania i usuwania rozproszonych zakleszczeń.
32.	Kryteria globalnej uszeregowalności konfliktowej.
Zaawansowana eksploracja danych	
33.	PageRank bez skoków, PageRank i Personalized PageRank: wyjaśnij różnice w działaniu algorytmów analizy linków.
34.	Opisz jak działa algorytm TF-IDF.
35.	Systemy rekomendacyjne oparte na użytkownikach i produktach: mechanizm działania, wady i zalety.
36.	Algorytm LSH i jego zastosowania.
Zaawansowane technologie przetwarzania danych	
37.	Standardy XPath, XSL i XQuery.
38.	Relacje przestrzenne w zapytaniach do przestrzennych baz danych, modele 9IM i DE-9IM.
39.	Standard SQL/MM: motywacje, koncepcja, zakres.
40.	Struktury fizyczne wspierające przeszukiwanie baz dokumentów tekstowych.
41.	Deskryptory wizualne wykorzystywane podczas wyszukiwania obrazów na podstawie ich zawartości (CBIR).
Modelowanie i analiza procesów biznesowych	
42.	Konstruktory modelu BPMN.
43.	Analiza własności procesów za pomocą sieci Petriego.
44.	Odkrywanie modelu procesów na podstawie zawartości logów.

Architektury zorientowane na usługi	
45.	Usługi sieciowe REST: definicja, komunikacja, dane.
46.	Założenia architektury mikrousługowej.
47.	Usługi sieciowe SOAP: definicja, komunikacja, dane.
48.	Konteneryzacja aplikacji, systemy zarządzania kontenerami.
49.	Założenia architektury SOA, rodzaje usług i ich charakterystyka.
Analiza i eksploracja sieci społecznościowych	
50.	Generatywne modele sieci złożonych.
51.	Modelowanie procesów dyfuzyjnych w sieciach.
52.	Metryki ważności wierzchołków i krawędzi w sieciach złożonych.
Zarządzanie bezpieczeństwem w systemach IT	
53.	Analiza ryzyka i metody postępowania z ryzykiem, polityki bezpieczeństwa.
54.	Modelowanie zagrożeń (STRIDE).
Zarządzanie projektami	
55.	Zarządzanie ryzykiem w projektach informatycznych.
56.	Zarządzanie zespołem.
57.	Zarządzanie jakością projektu (ang. design quality).