

Otázky k magisterské státní závěrečné zkoušce z předmětu
Databázové systémy a velká data
Obor: Datová věda
Akademický rok 2024/2025

1. **Big Data a NoSQL databáze** (Big Data – základní charakteristiky a 5V, NoSQL databáze - principy, NoSQL vs. relační databáze, modely dat, BASE, CAP teorém a problematika transakčního zpracování dat).
2. **NoSQL databáze klíč-hodnota** (pro vybraného zástupce uveďte jeho základní charakteristiky: architektura, možnosti distribuce dat, možnosti replikace a shardingu, principy, vhodné a nevhodné případy užití, možnosti zabezpečení).
3. **NoSQL databáze dokumentové** (pro vybraného zástupce uveďte jeho základní charakteristiky: architektura, možnosti distribuce dat, možnosti replikace a shardingu, principy, vhodné a nevhodné případy užití, možnosti zabezpečení).
4. **NoSQL databáze sloupcové** (pro vybraného zástupce uveďte jeho základní charakteristiky: architektura, možnosti distribuce dat, možnosti replikace a shardingu, principy, vhodné a nevhodné případy užití, možnosti zabezpečení).
5. **NoSQL databáze grafové** (pro vybraného zástupce uveďte jeho základní charakteristiky: architektura, možnosti distribuce dat, možnosti replikace a shardingu, principy, vhodné a nevhodné případy užití, možnosti zabezpečení).
6. **Distribuované databázové systémy** (výhody, nevýhody, základní funkce. Návrh distribuované databáze – fragmentace a alokace dat. Replikace dat a možnosti jejího využití).
7. **Business Intelligence** (základní pojmy a architektura BI, multidimenzionální modelování, OLAP, operace nad datovou kostkou, Self-Service BI, KPI, reporting).
8. **Datové sklady** (oblasti použití, porovnání analytických a transakčních databází – OLAP/OLTP, ETL/ELT procesy, datová kvalita, způsoby budování datových skladů, datový trh, bezpečnost).
9. **Formální modely distribuovaných systémů** (základní abstrakce distribuovaných systémů, architektury distribuovaných systémů, synchronní a asynchronní systémy).
10. **Protokoly a algoritmy pro řešení konsenzu** (Paxos, multi-Paxos a další algoritmy).
11. **Velká data v kontextu distribuovaných výpočtů** (klíčová řešení pro velká data - Hadoop (HDFS) a Spark, paralelní zpracování velkých datových souborů (MapReduce)).
12. **Ochrana dat a informací** (základní pojmy, rizika a protiopatření, zabezpečení přenosů, síťových řešení, zálohování a tisku, přístupová práva, možnosti ochrany HW, třídy a standardy bezpečnosti OS, firewall, bezpečnostní politika a funkce v organizaci bezpečnostní analýza, prognóza a projekt, rizika integrace a e-činnosti).
13. **Kryptologie** (Základní přístupy a pojmy kryptografický systém, klíč, časová a paměťová náročnost, kryptografický protokol, monoalfabetické a polyalfabetické šifry, symetrické a nesymetrické šifrování a moderní šifrovací systémy - principy, DES, RSA, atd.).
14. **Datové formáty v oblasti Big Data a NoSQL databází** (pro strukturovaná, nestrukturovaná a polostrukturovaná data a formáty pro specifické účely, výhody a nevýhody, příklady použití).

- 15. Bezpečnost a škálovatelnost NoSQL databází** (bezpečnostní výzvy v NoSQL databázích, škálovatelnost a dostupnost NoSQL databází, replikace a shardování v NoSQL databázích, případové studie škálování NoSQL řešení).
- 16. Real-time zpracování dat** (technologie pro streamové zpracování dat: Apache Kafka, Apache Flink, řešení pro zpracování dat v reálném čase, příklady aplikací real-time zpracování dat, integrace streamového zpracování s NoSQL databázími)