VŠB TECHNICKÁ UNIVERZITA OSTRAVA
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

# Key-value databázové systémy

Key-value database systems

Bc. Jan Jedlička

Bakalářská práce

Vedoucí práce: prof. Ing. Michal Krátký, Ph.D.

Ostrava, 2023

## Abstrakt

Cílem diplomové práce je popsat Key-value databázové systémy, ukázat výhodo těchto systémů a představit jedny z jejich významných představitelů. Součástí práce je návrh a implementace testovacího prostředí pro testování těchto systémů s ostatnímy SŘBD. Práce je zakončena vyhodnocením výsledků testů vybraných databázových systémů.

## Klíčová slova

NoSQL; Key-value databáze

## Abstract

The aim of the diploma thesis is to describe Key-value database systems, to show the advantage of these systems and to present some of their important representatives. Part of the work is the design and implementation of a test environment for testing these systems with other DBMS. The work is finished with an evaluation of the test results of selected database systems.

## Keywords

NoSQL; Key-value database

## Poděkování

Rád bych na tomto místě poděkoval vedoucímu diplomové práce panu prof. Ing. Michalovi Krátkému, za pravidelné konzultace a poskytnutí mnoha užitečných rad a nápadů pro řešení samotné práce.

# Obsah

**Seznam použitých symbolů a zkratek** **5**

**1 Úvod** **6**

**2 NoSQL Key-value databázové systémy** **7**

2.1 Amazon DynamoDB . . . . . 7
2.2 Oracle NoSql Database . . . . . 8
2.3 InfinityDB . . . . . 8
2.4 Redis . . . . . 9
2.5 Aerospike . . . . . 10
2.6 Oracle Berkeley DB . . . . . 11
2.7 Riak KV . . . . . 11
2.8 Voldemort . . . . . 12
2.9 Porovnání KV DB . . . . . 14
2.10 Nezmíněné významné KV DB v 2022 . . . . . 14

**3 Prostředí pro testování databázových systémů** **15**

**4 Vyhodnocení výsledků testů** **16**

**5 Závěr** **17**

**Literatura** **18**

# Seznam použitých zkratek a symbolů

| | | |
|---|---|---|
| NoSQL | – | No Structured Query Language |

# Kapitola 1

# Úvod

TODO úvod práce

# Kapitola 2

# NoSQL Key-value databázové systémy

TODO co jsou obecně no sql key value db systémy, výhody nevýhody obecně

## 2.1 Amazon DynamoDB

- největšá a nejpoužívánější Key Value DB
- serverless cloud databáze
- odezva microsekundy
- web tech, IoT, mobile, gaming
- plně spravovatelná multi master databáze
- odhadovatelný výkon a bezproblémová škálovatelnost
- unikátní primární klíč pro identifikaci jednotlivých záznamů v tabulce
- sekundární index pro lepší dotazovací flexibilitu
- primírní klíč je jako vstup do hash funkce, výsledný hash je fyzická pozice uloženého záznamu
- silná konzistence na čtení hodnot od poslední aktualizace
- atomic counters pro akutomatické změny hodnot číselných atributů
- TTL pro prošlé záznamy v tabulkách
- full backup pro archivaci dat
- automatická správa systému
- DynamoDB konzolové api pro správu db
- VPC pro soukromou komunikaci bez potřeby využití internetu

## 2.2 Oracle NoSql Database

- vhodná pro velké objemy dat a nízkou odezvu
- velký počet storage uzlů pro lepší výstup a kapacitu uložiště
- díky DB Java Edirion high-availability storage engine
- single master, multi replica - vysoká dostupnost a nízká chybovost
- při chybě na masteru je jedna z replik automaticky prohlášena za nový master
- grafová databáze
- itegrace v různých Oracle a Open Source aplikacích
- spolehlivá a flexibilní správa konfigurace skupiny uzlů
- škálovatelná se spolehlivostí pro poskytnutí správy všech dat pro zajištění spolehlivosti
- uzly a hrany v grafu reprezentují entity které vytvářejí vztahy a propojení
- sdílený systém, uniformně alokuje data okolo ostatních částí skupin
- storage uzly jsou replikované pro udržení konzistence
- obsahuje SQL Query s jazykem pro import, export a přenos dat mezi různými Oracle NoSQL databázemi
- Failover, SwitchOver, Bulk Get API, Off Heap Cache a podpora Big Data SQL a Apache Hive integration

## 2.3 InfinityDB

- Nestable Multi-value, může reprezentovat stromy, grafy, Key/Value mapy, dokumenty, velká řidká pole, tabulky
- ACID pro vlákna, ACD pro bulk operace
- jednoduché API, instantní produktuvita developerů
- dynamický pohled dotazů (set logic views, delta views, ranges)
- samostatná administrace, jeden soubor (bez konfiguračních souborů, logů, dočasných souiborů, upgrade skriptů a DBA)
- není potřeba dělat čištění junk souborů po operacích když zde nejsou žádné zanechýny

- runtime schema evolution (pro dopřednou a zpětnou kompatibilitu)
- all-Java db pro servery, procovní stanice, ruční zařízení
- protokol robustní vnitřní uložiště pro vytrvání na požádání nebo rozdělení cache na disk pro velké množství dat
- single data file, aktuální, bezpečný, korektní a využitelný pro každý případ (je designovaný pro použití jednoho souboru)
- bez zotavení na základě logu, restart a zotavení je vždy okamžité
- kombinace jednoho souboru a instantního zotavení nevyžaduje administraci db
- prostor pro uložení strukturovaných, polostrukturovaných a nestrukturovaných dat, tento jednoduchý model umožňuje ukládání stromů, grafů apod.
- možnost využít In-Memory-Only db která nechá všechna data v cachy, nebo naopak se data ukládají normálně do souboru, je možnost si vybrat bez změny kódu
- přístup k datům v cachy je plně více vláknový
- data která nejsou často využívaná jsou stránkována na disk

## 2.4 Redis

- in-memory uložiště pro datové struktury, využívaná jako db, cache nebo zprostředkovávatel zpráv
- můžeme jednou za čas ukládat pravidelně data na disk, nebo provádět logování při provádění operací (nemusíme nic ukládat a mít čistě in memory db)
- podpuruje datové struktury jako stringy, hashe, listy, množiny, bitmapy, hyperloglog a geospatial indexy
- Keyspace notifikace dovoluje klientům odebírat Pub/Sub kanály
- setřízené množiny pro vytváření indexů dle ID nebo jiného číselného atributu
- Geo APi pro dotazy na souřadnice nebo rádiusy
- Radius hashing ukládá data jako klíč a mapu
- Single-rooted replikovaný strom
- API pro populární používané jazyky

- obsahuje transakce, lua skripty a různé úrovně trvání na disku
- atomické operace (rozšíření stringu, přídání prvku do listu atp.)
- podpora trivial-to-setup master-slave asynchronního replikování
- rychlá neblokující se prvotní synchronizace, automatické znovu připojování se znovu synchronizací na netsplit
- skvělé využití pro Key = hash a Value = velký json objekt

## 2.5 Aerospike

- architektura hybrid memory
- internet scale, odhadovatelná a vysoká výkonost
- korektnost, silná konzistence, nízká cena, lineární škálovatelnost
- real-time rozhodování na velké, stále aktualizované db
- dynamická optimalizace pro optimální využití zdrojů
- server-side clustering
- bezpečnost na transportní vrstvě
- banking, telekominkace, adtech, gaming
- customer deployment s zero downtime
- transakce v milisekundách
- výběr in-memory storage (cache sessions) nebo SDD disk storage (pro trvanlivost dat) bez kompromisů ve výkonnosti
- silný quesry jazyk, vlastní vytvořitelné agregační funkce pomocí Lua jazyku (flexibilní pro agregační algoritmy)
- schema-less, sets/bins můžou být přidány za běhu (maximální flexibilnost pro aplikace)

## 2.6 Oracle Berkeley DB

- knihovna pro službu key-value db
- čistě in-memory db
- ultra-low odezva, škálovatelná, vysoce výkonná
- velice dostupná, tolerance chybovosti
- B-stromy, fronty, hash data indexy
- obnovitelné ACID transakce, více úrovní izolace (včetně MVCC)
- rozdělení dat dle key ranges
- možnost komprese záznamů
- jednoduché volání API pro přístup k datům a nastavení db
- stavební části pro db od lokálního uložiště po world-wide distribuovanou db (od KiloBytů po PetaByty)
- data uložená v XML, SQL (když není potřeba tak se využívá právě KeyValue uložiště), Java Objekty
- podpora moderních programovacích jazyků (C++, C#, Java, Python atp.)
- Single master, multi replica, vysoce dostupná konfigurovatelnost db
- repliky umožňují čtecí škálovatelnost, rychlý fail-over, hot-standby a další distribuované konfigurace dodávající podnikové prostředky v málém, vestavěném balíčku

## 2.7 Riak KV

- distibuovaná kv db, pokročilé lokální a multi-cluster replikování
- garance čtení a zápisu i při selhání HW nebo síťových oddílů
- bez konfliktní replikované datové typy (CRDTs), flags, registry, čitače, množiny a mapy
- konfigurace aktivního clusteru, poskytování dat pro klienty díky nízké latenci z nejbližšího data centra
- dostupné zóny, multi claster repliky a redundance dat v geografickém regionu
- flexibilní datový model bez pre definovaného schématu

- vylepšené logování chyb a reporty
- automatická komprese dat pomocí Snappy kompresní knihovny
- vhodná pro ukládání velkého množství nestrukturovaných dat
- pro big-data aplikace, ukládání dat z připojených zařáízení a replikování dat do okolí
- automatický distribuce dat skrz cluster, pro robustnost, výkon
- master-less architektura, vysoká dostupnost, téměř lineární škálovatelnost za využití snadného přidání HW kapacity bez nutnosti mnoha operací
- nízká latence, vhodná pro chat/messaging aplikace
- možnost zpracování dat pro získání užitečných závěrů a akceschopných informací
- design pro horizontálůní škálovatelnost s komoditním HW, jednoduché pro rozšíření objemu dat bez potřeby vytváření komplexního sdílení
- bez vnocování restrikcí na hodnoty, session data mohou být enkódována mnoha způsoby a nevyžadují změnu schématu
- během nejvyšší zátěže nezhoršuje zápis a horizontální škálovatelnost, uživatelé jsou obslouženi bez problémů
- dobrý pro soukromé, veřejné i hybridní cloud nasazování

## 2.8 Voldemort

- distribuovaná kv db založena na Amazon DynamoDB
- automatická replikace dat skrz více serverů
- automatické rozdělování dat mezi servery, každý server obsahuje pouze část z celkových dat
- nastavitelná konzistenčnoust
- transparentní ošetřování chyb serverů
- zapojitelný storage-engine (MySQL, Read-Only) a serializace (Java Serialization, Thrift, Avro)
- verzování dat pro maximální integritu i během poruch
- každý uzel je samostatný a nezávislý, žádný centrální řídící uzel nebo uzel řídící řešení chyb

- dobrá výkonost na jeden uzel, 10-20 tisíc operací za sekundu (1 op. za 50 mikro sekund) dle HW, sítě, systému disku atp.

- podpora zapojitelné strategie pro rozpoložení dat, pro možnost distribuce dat skrz data centra která jsou mezi sebou geologicky velice vzdálená

- využívá JMX pro zlepšení viditelnosti pro interní monitorování a validaci dat

- In-Memory caching pro eliminaci oddělených částí cache, jednoduché a rychlé in-memory tenstování (unit testy)

- horizontální škálování čtení i zápisu

- API rozhoduje o replikování a místě ukládání dat, různé strategie pro specifické aplikace

- široké možnosti pro klíče i hodnoty díky serializaci, listy a tuply s pojmenovanými poli

- JSON data model pro serializaci ale v kompaktním bytovém formátu, typová kontrola dat dle očekávaného schématu

- hashovatelné schéma, vyhledávání dle primárního klíče a možnost modifikace jednotlivých hodnot

- jednoduchá distribuce skrz stroje protože data mohou být rozdělována dle primárních klíčů

- dostupnost a bezpečnost jednotlivých oddílů při vysoké propustnosi

## 2.9 Porovnání KV DB

| Název | Správa | Škálovatelnost | Odezva | Zotavení |
|---|---|---|---|---|
| Amazon DynamoDB | automatická, plně spravovatelná | vysoká, horizontální | mikrosekundy | logy, záloha, automatické obnovení |
| Oracle NoSql DB | nízká | horizontální | milisekundy | replika je prohlášena za master |
| InfinityDB | jeden soubor se vším | nízká | milisekundy | bez logů, okamžité ale ztrácíme data |
| Redis | plná | horizontální | milisekundy | z logů (logování snižuje výkon) |
| Aerospike | plně spravovatelná | lineární | milisekundy | logy, záloha |
| Oracle Berkeley DB | velká | horizontální | mikrosekundy | repliy |
| Riak KV | vysoká | horizontální, téměř lineární | milisekundy | multi cluster repliky, logování |
| Voldemort | vysoká | horizontální | milisekundy | repliky |

## 2.10 Nezmíněné významné KV DB v 2022

- MongoDB
- Couchbase
- Azure Cosmos DB

# Kapitola 3

# Prostředí pro testování databázových systémů

TODO

# Kapitola 4

# Vyhodnocení výsledků testů

TODO

# Kapitola 5

# Závěr

TODO

# Literatura


1. *Top NoSQL Key Value store Databases: Predictiveanalyticstoday* [online]. 2022. [cit. 2022-11-13]. Dostupné z: `https://www.predictiveanalyticstoday.com/top-sql-key-value-store-databases/`.
2. *Best Document Databases: G2* [online]. 2022. [cit. 2022-11-13]. Dostupné z: `https://www.g2.com/categories/document-databases`.