|  |
| --- |
| OSTRAVSKÁ UNIVERZITA  PŘÍRODOVĚDECKÁ FAKULTA  KATEDRA INFORMATIKY A POČÍTAČŮ |
| NoSQL a SQL data modelování a technologie  DIPLOMOVÁ PRÁCE |
| Autor práce: Bc. Martin Rucký  Vedoucí práce: Ing. Zdeňka Telnarová, Ph.D. |
| 2020 |

|  |
| --- |
| UNIVERSITY OF OSTRAVA  FACULTY OF SCIENCE  DEPARTMENT INFORMATIC AND COMPUTERS |
| NoSQL and SQL data modeling and technology  THESIS |
| Author:  Bc. Martin Rucký  Supervisor:  Inf. Zdeňka Telnarová, Ph.D. |
| 2020 |

ABSTRAKT

Český text abstraktu

*Klíčová slova:*

*(klíčová slova vypsaná na řádku, oddělená od sebe čárkami)*

**ABSTRACT**

The text of the abstract.

*Keywords:*

čestné prohlášení

Já, níže podepsaný/á student/ka, tímto čestně prohlašuji, že text mnou odevzdané závěrečné práce v písemné podobě je totožný s textem závěrečné práce vloženým v databázi DIPL2.

Ostrava dne

………………………………

podpis studenta/ky

|  |
| --- |
| Poděkování |
| Prohlašuji, že předložená práce je mým původním autorským dílem, které jsem vypracoval/a samostatně. Veškerou literaturu a další zdroje, z nichž jsem při zpracování čerpal/a, v práci řádně cituji a jsou uvedeny v seznamu použité literatury.  V Ostravě dne . . . . . . . . . . . .  . . . . . . . . . . . . . . . . . .  (podpis) |

OBSAH

[ÚVOD 10](#_Toc12342371)

[1.1 O tématu NoSQL 11](#_Toc12342372)

[CÍLE PRÁCE 12](#_Toc12342373)

[2 Databáze 13](#_Toc12342374)

[2.1 Důležité pojmy 13](#_Toc12342375)

[2.1.1 Data 13](#_Toc12342376)

[2.1.2 Datové entity 14](#_Toc12342377)

[2.1.1 Tabulky 14](#_Toc12342378)

[2.1.2 Indexy 14](#_Toc12342379)

[2.1.3 Události 14](#_Toc12342380)

[2.1.4 Pohledy 14](#_Toc12342381)

[2.1.5 Atribut a záznam 14](#_Toc12342382)

[2.1.6 Primární a cizí klíč 16](#_Toc12342383)

[2.1 Databázové modely 16](#_Toc12342384)

[2.1.1 Relační databáze 16](#_Toc12342385)

[2.1.2 Objektové databáze 16](#_Toc12342386)

[2.1.3 Objektově relační databáze 16](#_Toc12342387)

[2.2 Vztahy mezi tabulkami 17](#_Toc12342388)

[3 NOSQL databáze 18](#_Toc12342389)

[3.1 Podnadpis 18](#_Toc12342390)

[3.1.1 Podpodnadpis 18](#_Toc12342391)

[4 Structured Query Language 19](#_Toc12342392)

[4.1 Návrh tabulek 19](#_Toc12342393)

[4.1.1 Příklad vytvoření tabulky 19](#_Toc12342394)

[4.2 Porovnání SQL a NoSQL 20](#_Toc12342395)

[4.2.1 Podpodnadpis 20](#_Toc12342396)

[5 NOSQL databázové servery 20](#_Toc12342397)

[6 Problémová doména 20](#_Toc12342398)

[ZÁVĚR 21](#_Toc12342399)

[RESUMÉ 22](#_Toc12342400)

[SUMMARY 23](#_Toc12342401)

[SEZNAM POUŽITÉ LITERATURY 24](#_Toc12342402)

[SEZNAM POUŽITÝCH SYMBOLŮ 25](#_Toc12342403)

[SEZNAM OBRÁZKŮ 26](#_Toc12342404)

[SEZNAM TABULEK 27](#_Toc12342405)

[SEZNAM PŘÍLOH 28](#_Toc12342406)

ÚVOD

Databáze je pojem, který dnes už všichni známe a využíváme. Používáme jej pro sběr dat, jejich ukládání, spravování a operace. Lidé v praktickém světě ji potřebují pro evidenci knih, lidí (klientů banky, pojišťovny atd.). Počty záznamů v těchto databázích můžou přesahovat miliony. Pro práci s daty vznikly databázové systémy, například SQL, které využíváme u relačních databází.

<https://theses.cz/id/1xkybg/BP.pdf>

Tyto databáze jsou založeny na tabulkách(relacích), což jsou dvourozměrné struktury, které obsahují řádky(záznamy) sloupce(atributy).

<https://cs.wikipedia.org/wiki/Rela%C4%8Dn%C3%AD_datab%C3%A1ze>

V současné době, ale narůstá zájem o jiný systém a to NoSQL. Důvodem je nejen rapidní nárůst dat, ale hlavně jejich struktura, např. záznamy se liší v počtu atributů.

<https://theses.cz/id/1xkybg/BP.pdf>

* 1. O tématu NoSQL

S neustále rostoucím zájmem o tento databázový systém vzrůstá dostupnost a počet jejich zástupců (MongoDB, Druid, Neo4J atd.). Výhodou také tohoto systému je, že většina nabízených zástupců jsou dostupná formou open source software.

[file:///C:/Users/Martin/Downloads/zaverecna\_prace%20(2).pdf](file:///C:\Users\Martin\Downloads\zaverecna_prace%20(2).pdf)

NoSQL je doslovně kombinací dvou slov: No a SQL. Je to tedy technologie, která stojí proti SQL. Zkratka může být matoucí a mezi lidmi neexistuje jednotný názor na to, co znamená. Nejvíce rozšířený je ovšem ten, který tvrdí, že jde o acronym „Not only SQL.“ Ať už zkratka znamená cokoliv, NoSQL je dnes zastřešující název pro všechny databáze, které nejdou cestou známého RDBMS (relační systém řízení báze dat), ale jdou svojí vlastní, často spojovanou s velkými objemy dat. (!!Přepsat!!)

<https://theses.cz/id/1xkybg/BP.pdf>

Tuto práci jsem si vybral z důvodu zájmu o databáze a rozšíření znalostí o této problematice. Svoji, bakalářkou práci jsem vedl na téma databázi (dotazovací jazyk SPARQL) a na univerzitě to patřilo k mým nejoblíbenějším předmětům.

<https://www.kiv.zcu.cz/studies/predmety/db2/cviceni-nosql.html>

<https://dspace5.zcu.cz/bitstream/11025/26327/1/Pokorny.pdf>

CÍLE PRÁCE

Hlavním cílem práce je popsat výhody a nevýhody SQL a NoSQL databází a poukázat na typy aplikací, které jsou vhodné spíše pro NoSQL databáze. V diplomové práci navrhneme konkrétní NoSQL databázovou technologii, kterou budeme demonstrovat na implementaci zvolené problémové domény.

Dalším cílem je charakterizace SQL a NoSQL. U NoSQL budeme popisovat nejméně tři NoSQL databázové servery s důrazem s jejich využitím v praxi.

Dalším dílčím cílem je u těchto tří databázových serverů představit příklady jejich vytvoření databáze a poukázat na rozdíly mezi sebou, kde, jaký typ databázového serveru využít, tak hlavně od SQL.

Pro praktickou část jsme si vybrali jednoho NoSQL zástupce z dokumentárních typů. Jednoho z nejvíce využívaného, který má také obrovskou komunitu. Naším posledním cílem je tedy vyvinout databázovou aplikaci na této NoSQL technologii.

1. Databáze

Je to systém, který pracuje s daty. Jeho úkolem je sbírat data a jejich následné zpracování. Databáze jsou s námi už dlouho, a to ve formě papírové, ve které se stále ještě využívá. Například u lékaře existují stále papírové kartotéky a tyto kartotéky obsahují data (informace o pacientech). Data musí také udržovat vazby mezi sebou a díky tomu některé lidi můžeme i začlenit do stejných skupin (řazení podle příjmení, věku). V databázi lékárny, ale nestačí pouze mít pouze data(informace) o pacientech, je zde potřeba zavézt také sestřičku a doktora, kteří s těmito informacemi pracují.

To byl tedy základ v papírově formě, se kterým člověk pracuje dlouho a zná ho velice dobře. Teď začneme se základy databází v počítačích.

<https://www.adaptic.cz/znalosti/slovnicek/databaze/>

Databáze neboli datové základny jsou místo, kde jsou ukládány určitým způsobem organizované a strukturované informace. Přístup k těmto informacím obstarává systém, kterému se říká SŘBD (Systém Řízení Báze Dat).

<http://www.databaze.chytrak.cz/>

* 1. Důležité pojmy
     1. Data

Data, jsme si již vysvětlili na příkladu v kapitole 2, ale je možné mít data ve více formátech např. textové, zvukové, grafické atd. Data v textové formě nejsou pouze texty, jako takové, ale máme různé typy.

* Textový typ – znakový řetězec, který má zpravidla maximální délku znaků 255
* Číselný typ – pro uložení celých a reálných čísel s pevnou, tak i s plovoucí dese tinnou čárkou
* Logický typ – pro uložení logických hodnot True/False (Pravda/Nepravda)
* Datumový typ – pro uložení datumu a datumu s časem
  + 1. Datové entity

Datové entity, objekty, které tvoří databázi (tabulky, indexy, události, pohledy atd.).

* + 1. Tabulky

Tabulky si lze představit jako běžné dvourozměrnou tabulky, která mají pevně daný počet a význam jednotlivých položek, které tvoří sloupce tabulek. Do tabulky jsou ukládány záznamy ve formě řádků.

* + 1. Indexy

Indexy jsou databázová konstrukce, sloužící ke zrychlení vyhledávacích a dotazovacích procesů v databázi, definování unikátní hodnoty sloupce tabulky nebo fulltextového vyhledávání.

* + 1. Události

Události(eventy) jsou spouštěny v určitý (uživatelem definovaný) datum a čas nebo opakovaně s definovatelnou periodou.

* + 1. Pohledy

Pohledy(View) jsou objekty, který uživateli poskytují data ve stejné podobě jako tabulka. Na rozdíl od tabulky, kde jsou data přímo uložena, obsahuje pohled pouze předpis, jakým způsobem mají být data získána z tabulek a jiných pohledů.

<https://cs.wikipedia.org/wiki/Datab%C3%A1ze>

* + 1. Atribut a záznam

Atribut, je jednotlivý sloupec v tabulce (např. jméno, příjmení, adresa, rodné číslo) a data jsou jejich jednotlivé záznamy.

Záznamem(datem) označujeme jednotlivý řádek v tabulce (např. Petr Novák; Palackého 65, Turnov; 8970612/3615)

* + 1. Primární a cizí klíč

Primární klíč**,** je atribut, jehož hodnota je pro každý záznam jedinečná (vhodným primárním klíčem je např. rodné číslo, neboť to má každý občan České republiky jedinečné)

Cizí klíč slouží pro vyjádření vztahů (relací) mezi databázovými tabulkami. Umožňuje nám identifikovat, které záznamy z různých tabulek spolu navzájem souvisí.

<http://www.databaze.chytrak.cz/>

* 1. Databázové modely

Z hlediska způsobu ukládání dat a vazeb mezi nimi můžeme rozdělit databáze do základních typů nejpoužívanější jsou:

* + 1. Relační databáze

Relační databáze, je databáze založená na relačním modelu. Často se tímto pojmem označuje nejen databáze samotná, ale i její konkrétní softwarové řešení.

Relační databáze je založena na tabulkách, jejichž řádky obvykle chápeme jako záznamy a eventuálně některé sloupce v nich (tzv. cizí klíče) chápeme tak, že uchovávají informace o relacích mezi jednotlivými záznamy v matematickém slova smyslu.

* + 1. Objektové databáze

Objektové databáze jsou systémy správy databázi, ve kterých informace jsou představeny formou objektů stejně, jako v objektově orientovaných programovacích jazycích.

* + 1. Objektově relační databáze

<https://cs.wikipedia.org/wiki/Datab%C3%A1ze>

* 1. Vztahy mezi tabulkami

Relace(vztahy) slouží ke svázání dat, která spolu souvisejí a jsou umístěny v různých databázových tabulkách. V zásadě rozlišujeme čtyři typy vztahů.

* mezi daty v tabulkách není žádná spojitost, proto nedefinujeme žádný vztah
* 1:1 (záznamu odpovídá právě jeden záznam v jiné databázové tabulce a naopak)
* 1:N (přiřazuje jednomu záznamu více záznamů z jiné tabulky)
* jedná se o nejpoužívanější typ relace, jelikož odpovídá mnoha situacím v reálném životě
* M:N (umožňuje několika záznamům z jedné tabulky přiřadit několik záznamů z tabulky druhé)
* tento vztah bývá z praktických důvodů nejčastěji realizován kombinací dvou vztahů 1:N a 1:M, které ukazují do pomocné, tzv. vazební tabulky složené z kombinace obou použitých klíčů

<http://www.databaze.chytrak.cz/>

<https://cs.wikipedia.org/wiki/Datab%C3%A1ze>

1. NOSQL databáze

Text

* 1. Podnadpis

Text

* + 1. Podpodnadpis

Text

1. Structured Query Language

Jazyk SQL je určený pracovat s relačními databázemi(RDBMS). Jde o strukturovaný dotazovací jazyk uzpůsobený pro práci s daty. V praxi dotazování vypadá následovně, zašleme databázi dotaz na určitá data a databáze nám vrátí výsledek požadovaných dat.

Příkazy jazyka SQL by se daly rozdělit do těchto skupin:

* DDL (data definition language) - příkazy patřící do této skupiny vytvářejí či upravují strukturu databáze (např. tabulky). Např.: CREATE, ALTER, DROP...
* DML (data manipulation language) - příkazy, které slouží k získávání, ukládání a mazání dat v databázi. Např.: SELECT, INSERT, UPDATE, DELETE...
* DCL (data control language) - příkazy pro správu uživatelských rolí a práv. Např.: GRANT, REVOKE...
* TCL (transactional control language) - příkazy pro správu databázových transakcí. Např.: BEGIN, COMMIT, ROLLBACK...
  1. Návrh tabulek

Každá tabulka musí mít primární klíč. Pro jedinečnou identifikaci záznamů v tabulce musíme dodržovat vytváření primárních klíčů ve všech tabulkách.

Neopisujeme data z jiných tabulek pro vyjádření vztahu. Nevytvářet tabulky zbytečně z mnoha atributů. Raději vytvořit novou tabulku, třeba pro data, která se hodně opakují. Např. u lékaře to může být krevní skupiny nebo léky.

* + 1. Příklad vytvoření tabulky

CREATE TABLE [Pacienti] (

[idPacienta] INT NOT NULL PRIMARY KEY,

[jmenoPacienta] VARCHAR(50) NOT NULL,

[prijmeniPacienta] VARCHAR(50) NOT NULL,

[pohlaviPacienta] BIT NOT NULL,

[narozeniPacienta] DATETIME NOT NULL,

[rodneCisloPacienta] INT NOT NULL

)

CREATE TABLE příkaz pro vytvoření tabulky Pacienti(název tbaulky). idPacient je nenulový primární klíč označen pomocí PRIMARY KEY, který má datový typ INT. Dále máme pět nenulových(NOT NULL) atributů z rozdílnými datovými typy(VARCHAR, INT, DATETIME, BIT).

* 1. Porovnání SQL a NoSQL

Text

* + 1. Podpodnadpis

Text

1. NOSQL databázové servery
2. Problémová doména

ZÁVĚR

RESUMÉ

SUMMARY

SEZNAM POUŽITÉ LITERATURY

1. **Příjmení, Jméno.** *Název knihy.* Město vydání: Vydavatelství, 2003. 123-4-56-789123-4.

2. **Příjmení1, Jméno1 a Příjmení2, Jméno2.** Název webové stránky. *Název webu.* [Online] Produkční společnost, 23. Září 2006. [Citace: 19. Září 2008.] http://www.urladresa.cz. 12-3456-789-12.

SEZNAM POUŽITÝCH SYMBOLŮ

|  |  |  |
| --- | --- | --- |
| ABC |  | Význam první zkratky. |
| B |  | Význam druhé zkratky. |
| C |  | Význam třetí zkratky. |
|  |  |  |

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

SEZNAM PŘÍLOH