

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Návrh distribuovaného prostředí pro práci s Big Daty
Jméno autora:	Bc. Alena Chernetskaya
Typ práce:	diplomová
Fakulta/ústav:	Fakulta dopravní (FD)
Katedra/ústav:	K614
Oponent práce:	Ing. Ondřej Doležal
Pracoviště oponenta práce:	ČVUT Fakulta dopravní

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání odpovídá svou náročností diplomové práci. Kromě samotné realizace a problematiky NoSQL databází autor také projevil schopnosti v oblasti předzpracování dat.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Autor splnil zadání, která si definoval v úvodu práce. Práce kromě samotné realizaci DB obsahuje i část zabývající se předzpracováním dat před importem do DB.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Autor na úvod práce provedl srovnání SQL a NoSQL databází, s tím že dále rozebral jednotlivé typy NoSQL databází včetně příkladu užití. Také uvedl do problematiky praktické části, kde v práci seznamuje čtenáře s povahou dat a nastiňuje základní požadavky na DB, na základě těchto požadavků provádí výběr. V požadavcích na DB mi však chybělo více požadavků na samotnou DB (bude třeba data agregovat?, jak a čím bude probíhat zpracování dat z db?). Autor řeší v práci tak povahu dat, požadavek na dostupnost a uložení dat.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Na základě práce nelze zpochybnit autorovu odbornou úroveň. V práci je patrné spojení teoretického úvodu s praktickou částí práce.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
V práci se objevují pravopisné chyby a chyby interpunkce, dále lze vytknout opakování některých slov. Samotné členění práce je v pořádku a mezi kapitolami je návaznost.	

Výběr zdrojů, korektnost citací	C - dobře
<i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	
K práci jsou použity relevantní zdroje, nelze se však ubránit dojmu, že některé části (např. 3.6.1. až 3.6.3) nejsou řádně očitovány.	

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Vložte komentář (nepovinné hodnocení).

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práce popisuje rozdíly SQL a NoSQL DB a zabývá se porovnáním jednotlivých typů NoSQL, následuje praktická část, kde autor na základě povahy dat a dalších požadavků vybírá správně řešení a ve virtualizovaném prostředí jej také realizuje.

V práci jsem však postrádal některé další požadavky na distribuované prostředí, především požadavky pro další systémy, které budou s tímto clusterem pracovat.

V textu je také rozpor ohledně specifikace dat v části 3.5 je zmíněno, že data jsou sbírána každých 5 minut z každého detektoru. V části 3.6.4 je však uvedeno 5 vteřin. Vytknout lze i občasné pravopisné chyby.

Výše uvedené výtky však nemají výrazný vliv na celkový dojem z práce i z toho důvodu, že se jedná o komplexnější a tedy náročnější téma, ve kterém navíc autor řešil i část předzpracování dat. Na závěr je nutné ocenit praktický přínos této práce a fakt že výsledky této práce lze tak použít v praxi.

Otázky:

- 1. Z jakého důvodu byla praktická část provedena na 1 virtuálním stroji a ne na 3, aby to lépe demonstrovalo produkční nasazení?*
- 2. Jaký je rozdíl mezi bucketingem a partitioningem ?*
- 3. Jakým způsobem probíhá přidání dalšího nodu do clusteru ?*

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře.**

Datum: 31.8.2021

Podpis: