

## I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Big Data ETL pro bankovní data
Jméno autora:	Bc. Ondřej Slavíček
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	RNDr. Martin Svoboda, Ph.D.
Pracoviště oponenta práce:	Katedra počítačů

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	lehčí
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem hodnocené diplomové práce mělo být srovnání existujících ETL nástrojů pro zpracování Big Data v oblasti bankovního sektoru, a to s ohledem na jejich vlastnosti, identifikované výhody a nevýhody, ale také na základě jejich experimentálního porovnání nejenom z pohledu dosažené výkonnosti nebo funkčnosti. Jelikož práce neměla směřovat k návrhu nového přístupu nebo alespoň řešení nedostatků těch stávajících, ale jen v jejich uceleném porovnání, považuji zadaný úkol za lehký.	

Splnění zadání	nesplněno
<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>	
Zadání práce se na první pohled jeví jako splněné, text práce do jisté míry obsahuje všechny aspekty v zadání zmíněné, dle mého názoru ale vždy v minimalistickém pojetí, a tedy celkově v nedostatečném rozsahu. Experimentální porovnání založené na časové náročnosti běhu jednoho jediného konkrétního agregačního výpočtu (navíc triviálního) nad jedním konkrétním schématem dat (opět velmi jednoduchým) nelze považovat za dostatečné, subjektivní časovou náročnost instalace nástroje či seznámení se s danou technologií samotným autorem už vůbec ne. Zadání proto považuji za nesplněné; už jen proto, že z uvedeného popisu ETL nástrojů a provedených experimentů nelze činit žádné obecně platné závěry.	

Zvolený postup řešení	nesprávný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Existující ETL řešení byla navíc zmíněna jen dvě (pomineme-li Map Reduce jako základní technologii), a to jen ve velmi stručném výčtu bez jakýchkoli hlubších nebo ucelených informací. Nelze tedy mluvit o žádné kritické analýze, ve které by se např. systematickou diskuzí nejprve stanovila srovnávací kritéria, očekávaný rozsah funkčnosti nebo požadované vlastnosti nejenom na infrastrukturu, škálovatelnost či konzistenci (tedy klíčových aspektů jakéhokoli zpracovávání Big Data). Diskutabilní je navíc už jen fakt, že oba nástroje potřebují data z HDFS clusteru stahovat a výpočty realizovat lokálně, což je pochopitelně proti základním principům Big Data (data tak velká, že je jednoduše není možné ukládat ani zpracovávat na jednom uzlu, ať už jakkoli supervýkonném serveru).	

Odborná úroveň	D - uspokojivě
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Odborná úroveň práce je nízká, autor neprokázal získání hlubších znalostí ve zkoumané problematice ani schopnost řešení obtížnějšího problému.	

**Formální a jazyková úroveň, rozsah práce**

**E - dostatečně**

*Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.*

Rozsah práce je na první pohled uspokojivý, ve skutečnosti ale vlastní text od úvodu až po závěr má jen 55 stran. Hlavní problém však spočívá v tom, že celých 40 stran je věnováno popisu obecných vlastností Big Data, principů ETL či již zmíněných konkrétních nástrojů včetně Apache Hadoop, zatímco vlastnímu řešení je věnováno jen stran 15. První část by navíc šlo z celé práce prakticky v celém rozsahu vypustit, protože jde jen o základní znalosti na úrovni magisterského studia, do značné míry pokryté dokonce jen úvodní motivační přednáškou předmětu Databázové systémy 2. Po stylistické stránce je vlastní text práce napsán kvalitně, sazba je také převážně dobrá. Text však obsahuje velké množství nejrůznějších chyb, a to včetně gramatických. Celkově proto rozsah práce považuji za nedostatečný, formální a jazykovou úroveň za dobrou.

**Výběr zdrojů, korektnost citací**

**D - uspokojivě**

*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.*

Počet uvedených zdrojů je značný, to potvrzuje i přehled autora o základních aspektech řešené oblasti Big Data. Bohužel ale nejsou citovány prakticky žádné vědecké články nebo zdroje obdobné odborné kvality.

**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.*

Práce se v podstatě omezuje jen na uživatelské experimentování s vybranými technologiemi. Složitostí a rozsahem ne nepodobným jednotlivým domácím úkolům ze zmíněného předmětu DS2. Jak již bylo řečeno, autorem uvedené závěry nelze považovat za obecně platné, celkové přínosy práce jsou pak minimální.

**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.*

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **F - nedostatečně**.

Datum: 7.6.2018

Podpis: