

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	<b>Big Data ETL pro bankovní data</b>
<b>Jméno autora:</b>	<b>Bc. Ondřej Slaviček</b>
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Vedoucí práce:</b>	Ing. Martin Bém
<b>Pracoviště vedoucího práce:</b>	Adastra, s.r.o.

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání lze považovat za náročnější a vysoce aktuální, protože praktická řešení z dílen velkých technologických dodavatelů i open source projektů přichází s velkými přísliby, ale uvedený problém prostě neřeší nijak uspokojivě pro velké implementace, kterou jsou alfou a omegou pro prostředí Big Dat. Stále je potřeba vyřešit mnoho dílčích problémů typu volba technologie nebo governance, které jsou v prostředí Big Dat mnohem kritičtější než v případě klasické datové integrace s ohledem na kvalitu služeb a celkové náklady. Zvolená aplikační oblast bankovníctví je velmi zajímavá, protože banky v současnosti začínají více využívat výhod Big Data řešení.	

<b>Splnění zadání</b>	<b>splněno</b>
<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>	
Předložená práce splňuje zadání v definovaném rozsahu. V rámci zadání bylo posouzeny dvě možné varianty řešení dávkové zpracování v podobě klasického ETL zpracování v Big Data prostředí (řešení od společností Hitachi/Pentaho, Talend) a nativního Big Data ETL (MapReduce). Závěry potvrzují poznatky z naší praxe, kdy výkon ETL nástrojů stále nedostihuje výkon nativně implementovaných ETL přímo v Hadoopu. Závěry také úspěšně odpovídají i na některé doposud otevřené otázky, na které jsme v praxi narazili.	

<b>Zvolený postup řešení</b>	<b>správný</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup praktického ověření odpovídá povaze problému a současnému stavu poznání. Zajímavé by bylo ověření dopadu horizontálního škálování Hadoop clusteru, avšak rozsah tohoto problému zdaleka přesahuje obvyklý rámec diplomové práce z důvodu velké pracnosti a náročnosti na hardwarové prostředky.	

<b>Odborná úroveň</b>	<b>A - výborně</b>
<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 výborná a velmi přesně vystihuje aktuální problematiku vymezení jejich Big Dat a jejich dávkového zpracování dat. Obecně se problematice Big Data ETL věnuje málo pozornosti na teoretické úrovni, takže zde lze nalézt další přínos této práce. Velmi často se bagatelizuje reálný vliv CAP teorému na zpracování dat pro analytické účely, který je také analyzován. Dále na praktické úrovni v důsledku technické složitosti existuje pouze několika málo použitelných technických řešení pro Big Data ETL, které z valné většiny práce pokrývá.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>B - velmi dobře</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Typografická a jazyková úroveň je velmi dobrá a nemám k ní zásadních výhrad.	

<b>Výběr zdrojů, korektnost citací</b>	<b>A - výborně</b>
--	--------------------

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.

Aktivita studenta při zpracování práce je příkladná. Využití je studijních materiálů a pramenů je velmi dobré a velmi aktuální. Na poli Big Dat je poměrně velké množství literatury buď úzce zaměřené manažersky na praktické aplikace nebo teoreticky řešící dílčí specifické technologické aspekty. Práce našla správnou kombinaci obou těchto směrů. Citace jsou zvoleny vhodně. Převzaté prvky lze jasně odlišit od vlastních přínosů.

### 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.

Potvrzené praktické výsledky ilustrují současné obecné snahy o nalezení jednodušší nástrojů pro dávkové zpracování dat Big Dat. Volba nástrojů odpovídá zájmu reálných komerčních subjektů, které aktivně řeší problematiku Big Data ETL. Velmi hezké je řešení generátoru dat, který umožňuje generovat statisticky různě velké datové soubory a výrazně zjednodušuje výslednou interpretaci dosažených výsledků.

### 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.

Mé celkové hodnocení nejvíce ovlivnilo správné a racionální pojetí ETL v oblasti Big Data, které je v praxi často bagatelizováno do úloh programování MapReduce, což je z pohledu klasické datové integrace bohužel návrat o cca. 30 let zpátky a popření velkého množství poznatků z abstrakce a automatizace tohoto typu úloh nejen z pohledu vývoje. Za zásadní závěr považuji za poukázání na nutnost volby správného řešení ETL dle povahy řešené úlohy, kdy i méně výkonné řešení může vhodnější než čisté řešení v podobě nativního Big Data ETL (MapReduce).

#### Otázky:

Jaký by se teoreticky chovalo implementované řešení po migraci na větší cluster (tzn. po provedení horizontálního škálování)?

Nabízí se pro nativní Big ETL (ETL běžící přímo pomocí Hadoop technologií) i jiné technologie než klasický MapReduce?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 4.6.2018

Podpis: