

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Student: **Bc. Miroslav Kvasnica** Vedoucí práce: **prof. Ing. F. Schauer, DrSc.**

Studijní program: **Inženýrská informatika**

Studijní obor: **Informační technologie**

Akademický rok: **2016/2017**

Téma diplomové práce: **Nová generace datového skladu reklamního systému Sklik.cz**

Hodnocení práce:

	A	B	C	D	E	F
	Hodnocení: A – nejlepší; F - nevyhovující					
1. Obtížnost zadaného úkolu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Splnění všech bodů zadání	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Práce s literaturou a její citace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Úroveň jazykového zpracování	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Formální zpracování – celkový dojem	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Logické členění práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Vhodnost zvolené metody řešení	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Kvalita zpracování teoretické části	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Kvalita zpracování praktické části	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Výsledky a jejich prezentace	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Závěry práce a jejich formulace	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Přínos práce a její využití	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Spolupráce autora s vedoucím práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Výsledek kontroly plagiátorství:

Výsledek souběhu s jinými pracemi je negativní. Vyjádření vedoucího diplomové práci příkládám v dalším.

Celkové hodnocení práce:

Výsledná známka není průměrem výše uvedených hodnocení. Znamku uvede vedoucí dle svého uvážení dle klasifikační stupnice ECTS:

A – výborně, B – velmi dobře, C – dobře, D – uspokojivě, E – dostatečně, F – nedostatečně.

Stupeň F znamená též „nedoporučuji práci k obhajobě“.

Předloženou diplomovou práci doporučuji k obhajobě a navrhuji hodnocení

A - výborně.

V případě hodnocení stupněm „F – nedostatečně“ uveďte do připomínek a slovního vyjádření hlavní nedostatky práce a důvody tohoto hodnocení.

Další připomínky, vyjádření, náměty k obhajobě práce (možno pokračovat i na další stránce):

Diplomovou práci hodnotím zejména z hlediska okamžitého využití v reálném systému Seznam.cz. Staré řešení interního datového skladu bylo založeno na jedné instanci MySQL databáze. Tento datový sklad byl určen pro interní analytiku a globální statistiky (Business Intelligence). V této databázi celých pět let rostla data, až se rychlost odezvy některých analytických dotazů stala neúnosná a dotazy mnohdy nedoběhly. MySQL sama o sobě neposkytuje žádný způsob škálování

dat a jako "řádkově orientovaný engine" není příliš vhodná pro stavbu datových skladů a OLAP. Celý databázový ekosystém (data lake) a ETL proces (Extract, Transform, Load) se opírá o technologie Hadoopu a proto nové řešení přibližuje datový sklad těmto technologiím. Nový datový sklad je implementován za pomoci databázového management systému Cludera Impala. Ten funguje nad HDFS (Hadoop Distributed File System) a dokáže využívat vlastnosti Hadoop ekosystému. Impala umožňuje snadné lineární škálování dat a je tak připravena na stovky T B (terabajtů) dat. Impala umožňuje provádět velice náročné analytické dotazy. Je plně integrována do Hadoop technologií a změny v datech je možné provádět pomocí distribuovaných frameworků jako je Yarn nebo Spark, což umožňuje analyzovat data, které produkuje reklamní systém Seznam.cz.

Datum 29. 5. 2017

