

## POSUDEK VEDOUCÍHO BAKALÁŘSKÉ PRÁCE

**Jméno studenta:** Bára Smolová

**Název bakalářské práce:**

**Vytěžování a vizualizace nestrukturovaných a semistrukturovaných dat**

**Cíl práce:** Charakteristika a srovnání vybraných technologických platforem pro vytěžování a vizualizaci nestrukturovaných a semistrukturovaných dat

**Jméno vedoucího bakalářské práce:** Ing. Vladimíra Zádová, Ph.D.

	Výborně	Velmi dobře	Dobře	Neprospěl
<b>I. Hodnocení zpracování tématu diplomantem:</b>				
Splnění cíle práce	X			
Volba metod a jejich aplikace při zpracování tématu	X			
Hloubka provedené analýzy	X			
<b>II. Hodnocení struktury a obsahu práce:</b>				
Přehlednost a logická stavba (struktura) práce	X			
Aktuálnost a vhodnost použitých pramenů	X			
Schopnost bakaláře zpracovat získané podklady	X			
Přiměřenost a srozumitelnost závěrů práce	X			
Formulování vlastních názorů studenta	X			
<b>III. Hodnocení formy a stylu práce:</b>				
Formální úprava práce (text, tabulky, grafy)	X			
Stylistická úroveň práce	X			
Práce s českou literaturou včetně odkazů a citací		X		
Práce se zahraniční literaturou včetně odkazů a citací		X		

**Vyjádření minimálně v rozsahu 10 řádků k bakalářské práci z hlediska splnění jejích cílů, využití metod řešení a návrhů opatření včetně formální úpravy, práce s literaturou a její citace (uved'te na druhou stranu posudku).**

**Otázky k obhajobě bakalářské práce:**

- **Srovnání platforem Splunk a Colix bylo provedeno pro bankovní sektor. Do jaké míry ovlivnila tato volba výsledek zpracování?**
- **Bude probíhat testování srovnávaných platforem na jiných typech dat?**

**Práci doporučuji - nedoporučuji\* k obhajobě. (\*nehodící se škrtněte)**

**Bakalářskou práci navrhuji klasifikovat stupněm:      **výborně****

**Datum: 23. 5. 2016**

.....  
**Podpis vedoucího bakalářské práce**



## Slovní hodnocení práce:

Bakalářská práce se zabývá zpracováním nestrukturovaných a semistrukturovaných dat.

Hlavní cíl práce je zaměřen na technologické platformy pro vytěžování a vizualizaci těchto dat. Z hlediska funkcionality, zprostředkovatele, klientů a podporovaných operačních systémů byly charakterizovány čtyři platformy. Vedle platform byly charakterizovány i open source nástroje, které jsou využívány pro sběr, zpracování a vizualizaci dat v uvedených platformách. Pro podrobnější srovnání byly vybrány dvě platformy – Splunk a Colix. Jako kritéria srovnání si autorka práce stanovila výkon, funkcionalitu, náklady a návratnost investic, uživatelskou přívětivost a možný rozvoj aplikace.

Všechny zvolené platformy mohou zpracovávat velké objemy strukturovaných, semistrukturovaných a nestrukturovaných dat a to i v reálném čase, data mohou pocházet z různých zdrojů. Pro zpracování nestrukturovaných a semistrukturovaných dat je v platformách používána metoda strukturalizace. Pro srovnání platform byl vybrán bankovní sektor.

Dílními cíli práce jsou specifika zpracování nestrukturovaných a semistrukturovaných dat a zpracování dat na zvolené technologické platformě.

Práce se skládá ze sedmi kapitol, úvodu a závěru. První kapitola obsahuje zhodnocení současného stavu v oblasti zpracování nestrukturovaných a semistrukturovaných dat. Specifikům zpracování dat jsou věnovány následující dvě kapitoly. Zpracování dat je zaměřeno na dolování a vizualizaci textu. Hlavní cíl práce – charakteristika a srovnání technologických platform je obsaženo ve čtvrté a páté kapitole. Šestá kapitola je věnována zpracování nestrukturovaných dat na vybrané platformě.

V literatuře je uvedeno 29 titulů, z nichž je citováno, z toho je 18 odkazů na zahraniční literaturu. V bibliografii je uvedeno 5 titulů.

Po formální stránce lze vytknout dílní nedostatky - chybně uvedené odkazy na číslo obr. v textu (od obr. 24), ne vždy správně uvedené citace, v seznamu použité literatury schází nadpis u citované literatury. To však s ohledem na celkovou úroveň práce považuji za nepodstatné. Délka práce odpovídá požadavkům.

Práce splňuje zásady pro vypracování formulované v jejím zadání, lze ji hodnotit kladně jak z hlediska tématu, tak z hlediska stanovených cílů a jejich splnění.

Zvolené téma je aktuální, náročné jak z hlediska rozsahu tak i obsahu - zejména pro tento typ práce. Při zpracování tématu autorka prokázala schopnost zvládnout teorii i praxi. Doporučuji se tématu věnovat nadále a využít získané znalosti i zkušenosti v diplomové práci.

