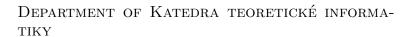
Sem vložte zadání Vaší práce.

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ





Diplomová práce

Analýza bezpečnostních rizik aplikací z logů v reálném čase

Bc. Vojtěch Krákora

Vedoucí práce: Pavel Pivoňka, GWCPM

Poděkování THANKS (remove entirely in case you do not with to thank anyone)

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval(a) samostatně a že jsem uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

Prague dne 2. dubna 2017

České vysoké učení technické v Praze Fakulta informačních technologií

© 2017 Vojtěch Krákora. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí, je nezbytný souhlas autora.

Odkaz na tuto práci

Krákora, Vojtěch. Analýza bezpečnostních rizik aplikací z logů v reálném čase. Diplomová práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2017.

Λ	bs	+ ~	1	.	ł
A	DS	Lr	a	K	L

V několika větách shrňte obsah a přínos této práce v českém jazyce.

Klíčová slova Replace with comma-separated list of keywords in Czech.

Abstract

Summarize the contents and contribution of your work in a few sentences in English language.

 ${\bf Keywords} \quad {\bf Replace \ with \ comma-separated \ list \ of \ keywords \ in \ English}.$

Obsah

In	trodi	action	1
1	Úvo	d do problematiky	3
	1.1	Zjišťování bezpečnostních rizik	3
	1.2	Platforma Unify	3
	1.3	Clustering	4
	1.4	Outliner	5
	1.5	Text mining	6
2	Náv	rh řešení	7
	2.1	Architektura aplikace	7
	2.2	Microsoft Azure	7
	2.3	JBoss	8
	2.4	Logování Unify	9
	2.5	Ukládání dat	9
	2.6	Předzpracování dat	10
	2.7	Vytvoření vektoru	10
	2.8	Konstrukce clusteringu	11
	2.9	Konstrukce outliner	12
	2.10	Prezentace dat	13
	2.11	Využití dát systémy 3. stran	13
3	Rea	lizace	15
	3.1	Nutné přípravy pro jboss	15
	3.2	Vytvoření modelu na Azure	16
	3.3	MongoDB	17
	3.4	Čtení dat z logů	17
	3.5	Předzpracování a odeslení do Azure	18
	3.6	Uložení dat	19
	3.7	Napojení na google charts	20

4	Analýza a vyhodnocení dat	23
	4.1 Analýza K-Means	23
	4.2 Analýza Outliner	24
5	Závěr	27
Co	onclusion	29
\mathbf{Li}^{\cdot}	teratura	31
\mathbf{A}	Acronyms	33
В	Contents of enclosed CD	35

Seznam obrázků

1.1	High Level Desing architektura Unify	4
	High Level Desing architektura aplikace	
2.2	Clustering k-means v prostředí MS AZURE ML Studio	11
2.3	Detekce anomálií v prostředí MS AZURE ML Studio	12
3.1	Prediktivní model clusteringu v Azure	16
3.2	Prediktivní model v detekci anomálií v Azure	17
3.3	Logo MongoDB. [1]	17

Introduction

[Napsat max jednu stranku] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Úvod do problematiky

1.1 Zjišťování bezpečnostních rizik

[[Co to jsou bezp rizika]] [[Jak se zjišťují]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

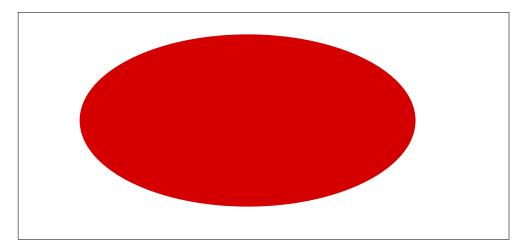
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.2 Platforma Unify

[[Představení Unify]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitu-

1. Úvod do problematiky

din. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Obrázek 1.1: High Level Desing architektura Unify.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.3 Clustering

[[Proč použít clustering]] [[Druhy clusteringu]] [[Výpočet vzdáleností]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus

velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.4 Outliner

[Proč použít Outliner]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

semper.

1.5 Text mining

[[Proč text mining]] [[Princip text miningu]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Návrh řešení

[[Nic moc tento odstavec]] V této kapitole se zabývám principy, technologiemi algoritmy, které jsem se rozhodl použít, k tomu abych splnil cíle této práce. Tedy vytvoření aplikace, jenž umožní sledovat bezpečností rizika v reálném čase.

2.1 Architektura aplikace

[[Pozor na duplici s odstavcem v úrovni nad]] [[Připsat, že to běží nad jboss]] Požadavkem na aplikaci je, aby požadavky zpracovávala a predikovala v reálném časem. Proto je princip položen na čtení požadavků proudících přes platformu. Jejich předzpracování a odeslání do cloudového řešení Microsoft Azure (Více v sekci 2.2). Po přijetí odpovědi je výsledek uložen do databáze. Pro vizualizaci dat slouží REST API [2], které vypisuje předem definované informace ve formátu pro ideální zobrazení v Google Charts [3].

Pro snažší představu o architektuře aplikace poslouží obrázek 3.3, na kterém je vidět High Level Desing [4].

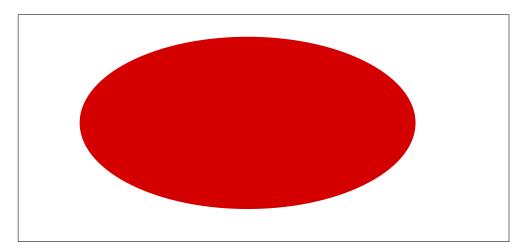
2.2 Microsoft Azure

Na integrační platformě Unify [5] je předpokládaný provoz 20 požadavků za vteřinu. Vzhledem k takto silnému provozu bude potřeba i přiměřeně velký výpočetní výkon.

Spolupráce se společností Miscrosof [6] mi umožnola jako řešení vyzkoušet její cloudové služby Microsfot Azure [7].

Microsoft Azure je sada integrovaných cloudových služeb. Azure nabízí cloudová řešení pro mnoho činností. Motivací k použití této služby k detekci bezpečnostích rizik je nástroj Microsoft Azure Machine Learning Studio [8].

Microsoft Azure Machine Learning Studio je plně cloudová služba, která umožňuje vytváření prediktivních modelů pro strojové učení [8]. Výhodou stu-



Obrázek 2.1: High Level Desing architektura aplikace.

dia je to, že veškerý výkon je rozprostřen vevnitř cloudu. Díky grafickému rozhraní je snadné vytvořit učící model, který je následně převeden do modelu prediktivního.

Aby měl prediktivní model smysl, je třeba mu poskytovat nějaká data, u kterých je predikce využita. K tomu se využívají webové služby. Prediktivní model se vystaví na specifické URL adrese. Zde je pak očekáván na vstupu konkrétní formát JSONu a služba vrací předem definovanou odpoveď se správnými parametry.

Výhodou využití cloudu je přenesení výpočetní zatěže mimo společnost. Naopak rizikem je problém s konektivitou, který může vytvořit výpadek služby a nebude tedy možné po tuto dobu predikovat rizika.

[[Více rozepsat rozdíl mezi učícím a prediktivním modelem]]

2.3 JBoss

Platforma Unify je postavená na aplikačním server JBoss AS 7 [5]. Z toho důvodu je třeba aby aplikace byla zcela kompatibilní.

[[Nají nějaké zdroje kde je popsáno o co vlastně jde]] Jboss AS je aplikační server pro Javu EE[9].

[[Popsat proč jboss]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollici-

tudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.4 Logování Unify

Platforma Unify loguje veškerý průběh do souborových audit logů. Protože celá integrace je založena na Javě, je využit logovací framework Log4j [10]. Knihovna Log4j umožňuje nastavit si pattern logování [11]. Na Unify je použit pattern [[Přidat pattern Log4j]].

Každý požadavek, který na platformu příde je zalogován do audit.logu. Zpráva je vždy na jedné řádce a zároveň na jedné řádce je povolena pouze jedna.

```
[[Ukázka logu]]
[[Popsat ukázku logu]]
```

Vzhledem k tomuto principu jsem se rozhodl přistoupit k tomu, že se vybrané logovací soubory budou kontinuálně číst, kde se bude ke každému řádku přistupovat jako k samostatné zprávě, která bude následně zpracována dál.

Díky této volbě nebude nutné nijak zasahovat do integrační platformy Unify a minimalizuje se tím riziko, jakéhokoliv nebezpečí ze strany mojí aplikace.

Unify využívá pro některé služby logování do Oracle databáze. Ale vzhledem k tomu, že jde pouze o několik málo určených služeb, připojení na databázy by tak kromě případných komplikací ani nepřineslo žádaný účinek.

2.5 Ukládání dat

[Porovnání SQL/NoSQL]] Rozhodl jsem se pro ukládání dat využít NoSQL databázi. Protože aplikace Unify momentálně ukládá veškerý svůj provoz do souborů (vyjímečně jsou některé konkrétní služby journalovány do DB), bude databáze využita i pro ukládání veškeré komunikace. To zpřístupní do budoucna snažší operování s jednotlivými zprávami. Popřípadě snažší zpětnou analýzu. [Schéma DB]]

Protože se bude ukládat veškerá komunikace, rozhodl jsem se v databázi mít uvedené následující informace:

[[Byt je to na pohled jasne, tak popsat co ktery param. je]]

- ObjectId automaticky generováno z MongoDB
- timestamp
- original-message
- normalized-message

- platformId
- assignment

2.6 Předzpracování dat

Pro snažší prácí s informacemi z logů jsem se rozhodl pro normalizaci jednotlivých požadavků. Normalizace je jeden z požadavků při zpracovávání textu [12].

V kapitole 2.4 jsou vidět informace, které se kromě samotné zprávy logují. V každé zprávě se objevuje takzvaná integrační hlavička. V té jsou základní údaje jako čas odeslání, jednoznačné identifikátory, zdrojové a cilové systémy. Položka jako je timestamp bude zpravidla pro každý požadavek jiná, stejně na tom logicku budou jednoznačné identifikátory. Z tohoto důvodu jsem se rozhodl zvolit jejich nahrazení.

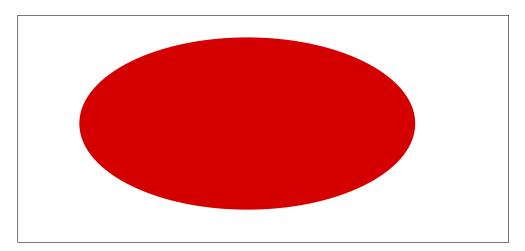
[[Přidat list toho jak nahrazuji čísla apod.]] [[Přidat ukázku]] [[Popsat celý proces, lowecase a podobně]]

2.7 Vytvoření vektoru

[Zde budu psát o tom, jak zprávu zpracouji pomocí algoritmu a převedu ji na číselný vektor]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.8 Konstrukce clusteringu

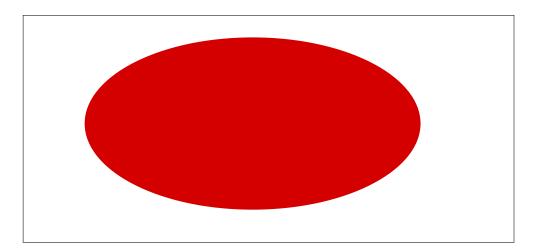
[[V této části bude trénovací algoritmus]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Obrázek 2.2: Clustering k-means v prostředí MS AZURE ML Studio.

2.9 Konstrukce outliner

[[Popsat]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Obrázek 2.3: Detekce anomálií v prostředí MS AZURE ML Studio.

2.10 Prezentace dat

Vzhledem k tomu, že by aplikace měla být schopna určovat bezpečnostní rizika, je třeba nějakým způsobem prezentovat její výstupy. Moninitoring na aplikaci Unify je momentálně postaven na tom, že konkrétní lidé hlídají logy a v případě vyskytu chyb, varování nebo jiné netypické události zjišťují co bylo příčinou.

Rozhodl jsem se tedy, že nejlepší bude grafické znázornění. Kromě údajů o tom, že byl zaznamenán požadavek, který je podezřelý budu grafy využívat i k prezentaci základního monitoringu.

Vzhledem k tomu, že se bude veškerá komunikace ukladát bude vhodné prezentovat například i kolik požadavků na jednotlivé komponentě proběhlo za poslední hodinu a podobně.

Společnost Cetin a.s. [13] ve které bude aplikace testována a jenž je uživatelem integrační platformy používá pro různá grafická znázornění grafy od Google Charts[3].

Tyto grafy jsou napsané v jazyce Javascript. Je tedy možné jejich umístění například na intranetové stránky, kde se vysoce postavení lidé společnosti vyznají lépe než v jednotlivých monitorovacích aplikacích.

Na tomto základě jsem se rozhodl vytvořit REST API [2], jenž budou Google Charts schopny snadno konzumovat a v případné jiné aplikace, které by stály o podobná data budou schopny se jim přizpůsobit.

2.11 Využití dát systémy 3. stran

Do budoucna je potřeba počítat s rozšířením monitoringu a je proto vhodné aplikaci připravit tak, aby její výsledky mohly být využity v aplikacích 3. stran.

Lze předpokládat, že k monirování bezpečnosti provozu budou použity systémy SIEM (Security Information and Event Management) [14]. SIEM funguje na principu, kdy zpracovává co nejvíce údajú, na jejichž základě pak rozeznává neočekávané situace a rizika [15].

Tím, že jsem se rozhodl data ukládat tak, jak uvádím v kapitole 2.5 bude libovolný SIEM po připojení do DB schopen získat jak originální zprávu, tak její normalizovanou verzi popřípadě i výsledek vyhodnocení mé aplikace.

Dále je možnost napojit SIEM i na REST API obdobně jako Google Charts v kapitole 2.10.

Realizace

[[Sem vepsat nějaký úvod k této kapitole]]

3.1 Nutné přípravy pro jboss

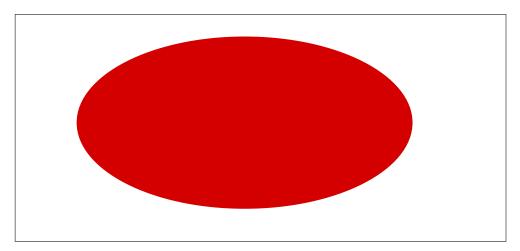
[Popsat jaké moduly] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.2 Vytvoření modelu na Azure

[[Ukázka URL]] [[Ukázka Vstupního JSONu a výstupního]] [[V této části bude Prediktivní algoritmus]]

3.2.1 Clustering v Azure

[Popsat prediktivní experiment] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

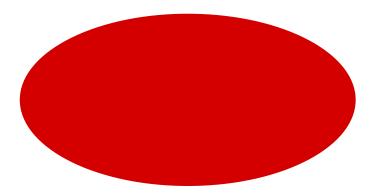


Obrázek 3.1: Prediktivní model clusteringu v Azure.

3.2.2 Detekce anomálií v Azure

[[Popsat prediktivní experiment]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus.

Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Obrázek 3.2: Prediktivní model v detekci anomálií v Azure.

3.3 MongoDB

Pro potřeby aplikace je nutné v co nejrychlejším možném čase ukládat jednotlivé requesty. Při ohromném provoze, který se na integrační platformě vyskytuje to je nutná podmínka pro to, aby bylo možné v reálném čase jednotlivé požadavky zpracovávat.



Obrázek 3.3: Logo MongoDB. [1]

Rozhodl jsem se za tímto účelem využít MongoDB [1], protože se očekává, že bude třeba ukládat 30 záznamů za sekundu, zpravidla bude docházet k více zapisům než čtením.

3.4 Čtení dat z logů

[[Popis]] Při návrhu zisku jednotlivých požadavků z integrační platformy jsem vycházel z toho, že nový aplikace musí minimálně, či spíše vůbec nezatěžovat

Unify [5]. Vzhledem k tomu, že přes integraci proudí veškerý provoz, je sama o sobě dosti vytížená a v případě, že by touto aplikací byl způsoben výpadek došlo by k silnému ztížení veškerých bussiness procesů, což si nelze dovolit.

Unify veškeré požadavky ukládá do logovacích souborů. Některé, převážně rizikové, služby se zároveň ukládají Oracle databáze. Ale vzhledem k tomu, že nejde o všechny dostupné služby rozhodl jsem se toho nevyužít.

Princip získání dat proudicích přes integrační platformu je založen na čtení jednotlivých logovacích souborů. Jako vhodný nástroj jsem vybral Java třídu Tailer z dostupné knihovny org.apache.commons.io [16].

Třída Tailer, po implementaci listeneru, se chová stejně jako linuxový příkaz tail [17]. Průběžně kontroluje čtený soubor a každou nově zapsanou řádku zpracovává.

Tímto řešením získáváme data z integrační platformy, aniž bychom jí zatěžovali.

3.5 Předzpracování a odeslení do Azure

Protože jsou data odesílána do cloudu, předzpracováváme je lokálně a přímo do Microsfot Azure odesíláme už jen identifikátor zprávy a vypočtený vektor.

Po přečtení zprávy z auditového logu Unify je zpráva předzpracována (2.6) a následně je z ní vytvořen vektor (2.7).

[[Popsat jak získám data]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat jak z dat udělám vektor]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat jak data odešlu]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat jak data přijmu]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.6 Uložení dat

[[Popsat v jakém stavu jsou data před ukládáním]] [[Popsat jak se data ukládají]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.7 Napojení na google charts

[Popsat jak vypadá javascript]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut portitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat obecné api, které google charts čeká]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat jak je vyřešené rest API]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque

Analýza a vyhodnocení dat

[[Sem vepsat nějaký úvod k této kapitole]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

[[Popsat systém, na který to bylo nasazené]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.1 Analýza K-Means

[[Popsat vstupní parametry]] [[Ukázat výsledky na Prod logu]] [[Dojít k závěru]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc,

molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.2 Analýza Outliner

[[Popsat vstupní parametry]] [[Ukázat výsledky na Prod logu]] [[Dojít k závěru]] Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan

semper.

Kapitola **5**

Závěr

Conclusion

Literatura

- [1] MongoDB. Available from: www.mongodb.com
- [2]
- [3] Google Charts. Available from: https://developers.google.com/chart/
- [4] Johnson, C. H. High Level Design Distributed Network Traffic Controller. 02 2005. Available from: https://people.ok.ubc.ca/rlawrenc/research/Students/CJ_05_Design.pdf
- [5] Unify integration platform. Available from: https://www.physter.com/unify/
- [6] Available from: https://www.microsoft.com/cs-cz/
- [7] Microsoft, . . Microsoft Azure. Available from: https://azure.microsoft.com/cs-cz/
- [8] Microsoft Azure Machine Learning Studio. Available from: https://studio.azureml.net/
- [9] DEPLOY ANYWHERE WITH RED HAT JBOSS ENTER-PRISE APPLICATION PLATFORM. Available from: https: //www.redhat.com/cms/managed-files/mi-deploy-anywhere-jbosseap-datasheet-inc04051031w-201605-en.pdf
- [10] Apache Log4j 2. Available from: https://logging.apache.org/log4j/2.x/
- [11] Class PatternLayout. Available from: https://logging.apache.org/log4j/1.2/apidocs/org/apache/log4j/PatternLayout.html
- [12] Sproat, R.; Bedrick, S. CS506/606: Txt Nrmlztn. 2011. Available from: http://www.csee.ogi.edu/~sproatr/Courses/TextNorm/

- [13] CETIN. Available from: https://www.cetin.cz/
- [14] Constantine, C. SIEM and Log Management Everything you need to know but were afraid to ask, Part 1. 2014. Available from: https://www.alienvault.com/blogs/security-essentials/everything-you-wanted-to-know-about-siem-and-log-management-but-were-afraid
- [15] Work?, H. D. S. Written by Colton Bachman. 2016. Available from: https://www.integritysrc.com/blog/313-how-does-siem-work
- [16] Tailer class. Available from: https://commons.apache.org/proper/commons-io/javadocs/api-2.4/org/apache/commons/io/input/Tailer.html
- [17] Tail. Available from: https://www.gnu.org/software/coreutils/manual/html_node/tail-invocation.html

PŘÍLOHA **A**

Acronyms

 ${\bf GUI}$ Graphical user interface

XML Extensible markup language

PŘÍLOHA **B**

Contents of enclosed CD

readme.txt	the file with CD contents description
exe	the directory with executables
src	the directory of source codes
T. Control of the Con	implementation sources
thesisthe dire	ectory of LATEX source codes of the thesis
text	the thesis text directory
thesis.pdf	the thesis text in PDF format
1	the thesis text in PS format