

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačního inženýrství**



## **Diplomová práce**

**Využití dat log souborů informačního systému  
k řídicím, analytickým a diagnostickým účelům**

**Bc. Josef Šimánek**

© 2012 ČZU v Praze

# ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Katedra informačního inženýrství

Provozně ekonomická fakulta

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Šimánek Josef

Informatika

Název práce

**Využití dat log souborů IS k řídicím, analytickým a diagnostickým účelům**

Anglický název

**Utilization of log files data of IS for management, analytical and diagnostic purposes**

### Cíle práce

Diplomová práce je tematicky zaměřena na problematiku využití dat log souborů IS k řídicím, analytickým a diagnostickým účelům. Cílem práce je:

- a) vymezit teoretické principy log souborů v IS cizinecké policie,
- b) zmapovat současnou úroveň využívání těchto souborů v cizinecké policii,
- c) navrhnout širší možnosti jejich využívání pro potřeby zefektivnění řízení cizinecké policie,
- d) navržené záležitosti ověřit a demonstrovat na funkčních modulech,
- f) ověřené záležitosti zobecnit pro další možná použití.

### Metodika

Použitá metodika řešené diplomové práce bude založena na studiu a analýze dostupných informačních zdrojů. Rovněž bude použity metody relační databázové technologie, SQL, regresní analýza a další vhodné statistické metody ( charakteristika polohy, variability, korelační a regresní analýza). Navrhované řešení bude realizováno formou praktického řešení, které bude respektovat identifikované požadavky na tato řešení a ověří navrhované záležitosti. Na podkladě syntézy teoretických poznatků a výsledků budou formulovány závěry této diplomové práce a následně zobecněny pro další možná použití.

### Harmonogram zpracování

Vymezení teoretických principů řešené problematiky, 1. zápočet z DP za 4.ročník: 06/2011-09/2011

Zmapování současné úrovně využívání log souborů v cizinecké policii, identifikování dalších  
potencionálních možností: 09/2011 - 11/2011

Navržení využití identifikovaných potencionálních možností využívání log souborů, 1. zápočet z DP  
za 5. ročník: 11/2011-01/2012

Ověření navržených záležitostí a jejich zobecnění, 2 zápočet z DP za 5. ročník: 01/2012 - 03/2012

**Rozsah textové části**

60 - 80 stran

**Klíčová slova**

Informační systém, informační zabezpečení, Log soubor, transformace, databázový systém, SQL

---

**Doporučené zdroje informací**

BOLDYŠ, P.: Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 1 – Citace: metodika a obecná pravidla. Verze 3.3. c1999-2004,  
SVATOŠOVÁ, L., KÁBA, B.: Statistické metody I, PEF ČZU, Praha, 2009. ISBN 978-80-213-1672-0  
TAYLOR, A. G.: SQL for dummies. 5th edition. Indianapolis Willey Publishing. Computer Press. 2001. ISBN 07645-4075-0

---

**Vedoucí práce**

Vostrovský Václav, doc. Ing., Ph.D.

**Termín odevzdání**

březen 2012

---

**prof. Ing. Ivan Vrana, DrSc.**

Vedoucí katedry

---

**prof. Ing. Jan Hron, DrSc., dr.h.c.**

Děkan fakulty

V Praze dne 23.6.2011

## Čestné prohlášení

Prohlašuji, že svou diplomovou práci **Využití log souborů informačního systému k řídicím, analytickým a diagnostickým účelům** jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 19 . března 2012

---

## **Poděkování**

Chtěl bych touto cestou poděkovat vedoucímu diplomové práce doc. Ing. Václavu Vostrovskému, Ph.D. za odborné vedení a cenné rady, díky kterým jsem tuto práci úspěšně dokončil.

Zároveň chci poděkovat rodině za podporu a zázemí, které mi poskytovala po celou dobu studia.

**Využití dat log souborů informačního systému  
k řídicím, analytickým a diagnostickým účelům**

---

**Utilization of log files data of information system  
for management, analytical and diagnostic purposes**

## **Souhrn**

Diplomová práce se zabývá problematikou využití dat z logovacích souborů informačního systému cizinecké policie, a to z hlediska řídicích, analytických a diagnostických účelů. Teoretická část popisuje současný stav využívání informací cizineckého informačního systému a historii nasazení výpočetní techniky u této složky. V další části je proveden rozbor dat, která shromažďují log soubory s návrhem jejich dalšího využití. Získaná data jsou nadále zpracována prostřednictvím funkčních modulů, postupů, výpočtů a grafů.

Na závěr práce je provedeno vyhodnocení navrhovaného řešení a posouzen jeho přínos pro budoucí nasazení v rutinní praxi.

## **Klíčová slova**

informační systém

informační zabezpečení

LOG soubor

transformace dat

databázový systém

SQL

## Overview

The diploma thesis solves the problem of application data from log files to the information system of the foreigner police in terms of control, analytic and diagnostic purposes. The theoretical part describes the current situation of usage information from foreigners' information system, and history of deployed computing technologies for this sector. In the next part is executed analysis of date from collected log files, suggest the next usage. The obtained dates are processed by functional modules, procedures, calculations and graphs.

The conclusion of this thesis is the suggested solution, and the evaluation of his benefit for putting into routine practice.

### Key Words:

Information System

informative safeguard

LOG files

data transformation

database system

SQL



## Obsah:

<b>1</b>	<b>ÚVOD</b> .....	<b>8</b>
<b>2</b>	<b>CÍL PRÁCE A METODIKA</b> .....	<b>11</b>
2.1	METODIKA .....	13
2.1.1	<i>Práce s relační databází</i> .....	13
2.1.2	<i>Využití statistických metod</i> .....	14
2.1.3	<i>Automatizace některých procesů</i> .....	15
2.1.4	<i>Práce s výstupy</i> .....	15
2.1.5	<i>Práce s textovým dokumentem</i> .....	16
<b>3</b>	<b>PŘEHLED ŘEŠENÉ PROBLEMATIKY</b> .....	<b>17</b>
3.1	OUTSOURCING, HISTORIE A UPLATNĚNÍ U CIZINECKÉ POLICIE.....	17
3.2	LOGOVACÍ SOUBOR .....	18
3.2.1	<i>Struktura logovacího zápisu</i> .....	20
3.2.2	<i>Formát zápisu logů</i> .....	20
3.3	CIZINECKÝ INFORMAČNÍ SYSTÉM (CIS) .....	22
3.3.1	<i>Skladba databáze programu CIS a vnější zdroje</i> .....	23
3.3.2	<i>Bezpečnostní LogSoubor v CIS</i> .....	25
3.3.3	<i>Rozdělení kompetencí při provozu CISu</i> .....	25
3.4	STRUKTURA BEZPEČNOSTNÍHO LOGSOUBORU CIS.....	27
3.4.1	<i>Struktura logu dotazu („D“):</i> .....	27
3.4.2	<i>Struktura logu odpovědi („O“):</i> .....	28
3.4.3	<i>Struktura logu příloh („P“):</i> .....	29
3.4.4	<i>Objem dat bezpečnostních LogSouborů CIS</i> .....	29
3.5	STRUKTURA RELAČNÍ DATABÁZE A VÝBĚR DAT Z LOGSOUBORU .....	30
3.5.1	<i>Jednotlivé položky relační tabulky</i> .....	30
3.6	TRANSFORMACE DAT Z LOGSOUBORU DO RELAČNÍ DATABÁZE .....	31
3.6.1	<i>Časová náročnost transformace dat</i> .....	33
<b>4</b>	<b>PRAKTICKÁ ČÁST</b> .....	<b>35</b>
4.1	VYUŽITÍ DAT Z LOGSOUBORU K ŘÍDÍCÍM ÚČELŮM .....	35
4.1.1	<i>Porovnání počtu cizinců a zatížení CIS podle spádových oblastí</i> .....	35
4.1.2	<i>Hodinové zatížení serverů dotazy v roce 2009 a 2010</i> .....	39
4.1.3	<i>Porovnání plnění pokynu podle oblastí</i> .....	43
4.2	VYUŽITÍ DAT Z LOGSOUBORU K STATISTICKÝM ÚČELŮM .....	46
4.2.1	<i>Základní statistické údaje z CIS</i> .....	46
4.2.2	<i>Počty pohybu uživatelů v programovém prostředí CIS</i> .....	49
4.2.3	<i>Porovnání časové náročnosti jednotlivých modulů</i> .....	52
4.2.4	<i>Základní statistické údaje o vygenerovaných odpovědích</i> .....	54
4.3	VYUŽITÍ DAT Z LOGSOUBORU K DIAGNOSTICKÝM ÚČELŮM.....	56
4.3.1	<i>Porovnání vytiženosti aplikačních serverů v hodinovém intervalu</i> .....	56
4.3.2	<i>Porovnání funkčnosti „BalancServeru“</i> .....	60
4.3.3	<i>Závislost počtu odpovědí na době zpracování dotazu</i> .....	62
<b>5</b>	<b>ZHODNOCENÍ A DOPORUČENÍ</b> .....	<b>68</b>
5.1	NASAZENÍ DO PRAXE .....	68
5.2	EKONOMICKÁ ZÁTĚŽ.....	69
5.3	DOPORUČENÍ.....	70
<b>6</b>	<b>ZÁVĚR</b> .....	<b>72</b>

<b>7</b>	<b>SEZNAM POUŽITÝCH ZDROJŮ .....</b>	<b>75</b>
<b>8</b>	<b>PŘÍLOHY .....</b>	<b>78</b>
8.1	PŘÍLOHA 1 .....	78
8.2	PŘÍLOHA 2 .....	80
8.3	PŘÍLOHA 3 .....	82
8.4	PŘÍLOHA 4 .....	84
8.5	PŘÍLOHA ČÍSLO 5 .....	85

## Seznam obrázků:

OBRÁZEK 1 - ZNÁZORNĚNÍ NAVRHOVANÉHO ŘEŠENÍ.....	11
OBRÁZEK 2 – ZNÁZORNĚNÍ ROZDĚLENÍ ODPOVĚDNOSTI ZA PROVOZ CIS .....	17
OBRÁZEK 3 – UKÁZKA GRAFICKÉHO PROSTŘEDÍ CIS S ROZSÁHLOU NABÍDKOU VYHLEDÁVACÍCH NÁSTROJŮ .....	23
OBRÁZEK 4 - ZNÁZORNĚNÍ VYUŽITÍ BALANCER SERVERU A UKLÁDÁNÍ DAT V CIS (VYTVOŘIL AUTOR).....	24
OBRÁZEK 5 – ZNÁZORNĚNÍ ROZDĚLENÍ KOMPETENCÍ PŘI PROVOZU CIS .....	26
OBRÁZEK 6 – GRAF POROVNÁNÍ POČTU DOTAZŮ S POČTY CIZINCŮ VE SPÁDOVÝCH OBLASTECH V PROCENTECH .....	37
OBRÁZEK 7 - GRAF POROVNÁNÍ POČTU DOTAZŮ JEDNOTLIVÝCH DNÍ V TÝDNU .....	38
OBRÁZEK 8 – GRAF POROVNÁNÍ HODINOVÉHO INTERVALU ZA MĚSÍC.....	41
OBRÁZEK 9 – GRAF POROVNÁNÍ HODINOVÉHO INTERVALU DOTAZŮ ZA TÝDEN .....	42
OBRÁZEK 10 - GRAF POROVNÁNÍ HODINOVÉHO INTERVALU DOTAZŮ ZA DEN .....	42
OBRÁZEK 11 - GRAF POROVNÁNÍ PLNĚNÍ POKYNU DLE SPÁDOVÝCH OBLASTÍ .....	45
OBRÁZEK 12 – GRAF POMOCNÝ .....	48
OBRÁZEK 13 - GRAF POROVNÁNÍ DENNÍHO VYUŽITÍ OBRAZOVEK .....	51
OBRÁZEK 14 - GRAF POROVNÁNÍ ČASOVĚ NEJNÁROČNĚJŠÍCH MODULŮ .....	53
OBRÁZEK 15 – GRAF POROVNÁNÍ POČTU DOTAZŮ DLE RYCHLOSTI ZPRACOVÁNÍ. ....	53
OBRÁZEK 16 - GRAF POROVNÁNÍ APLIKAČNÍCH SERVERŮ ZA MĚSÍC .....	57
OBRÁZEK 17 - GRAF POROVNÁNÍ APLIKAČNÍCH SERVERŮ ZA TÝDEN .....	58
OBRÁZEK 18 - GRAF POROVNÁNÍ APLIKAČNÍCH SERVERŮ ZA DEN .....	59
OBRÁZEK 19 - GRAF ROZLOŽENÍ ZATÍŽENOSTI POMOCÍ BALANCER SERVERU .....	61
OBRÁZEK 20 – GRAF POROVNÁNÍ ČETNOSTI DOTAZŮ V ZÁVISLOSTI NA INTERVALU POČTU DOTAZŮ.....	64
OBRÁZEK 21 - GRAF POROVNÁNÍ POČTU DOTAZŮ ZPRACOVANÉHO V ČASOVÉM LIMITU A NAD ČASOVÝ LIMIT .....	65
OBRÁZEK 22 - GRAF PROCENTUÁLNÍHO ZASTOUPENÍ NEPLNĚNÍ LIMITU DOBY ZPRACOVÁNÍ .....	66
OBRÁZEK 23 - GRAF ČETNOSTI DESETI NEJPOUŽÍVANĚJŠÍCH OBRAZOVEK V CIS.....	79
OBRÁZEK 24 - GRAF POROVNÁNÍ POČTU DOTAZŮ PODLE MODULU .....	81

## Seznam tabulek:

TABULKA 1 - STRUKTURA DOTAZU "D" .....	28
TABULKA 2 - STRUKTURA LOGU ODPOVĚDI "O" .....	28
TABULKA 3 - STRUKTURA LOGU PŘÍLOH "P" .....	29
TABULKA 4 - OBJEM DAT V BEZPEČNOSTNÍM LOGSOUBORU CIS .....	29
TABULKA 5 - POROVNÁNÍ SPÁDOVÝCH OBLASTÍ DLE CIZINCŮ A ZATÍŽENOSTI CIS .....	36
TABULKA 6 - HODINOVÉ ČETNOSTI DOTAZU NA SERVERY .....	40
TABULKA 7- POROVNÁNÍ PLNĚNÍ POKYNU ZADÁVÁNÍ DŮVODU DOTAZU DLE KRAJŮ .....	44
TABULKA 8 - DESET NEJZATÍŽENĚJŠÍCH HODIN DOTAZY OBOU SERVERŮ V ROCE 2009 A 2010 .....	47
TABULKA 9 - NEČINNOSTI NA APLIKAČNÍCH SERVERECH V PRŮBĚHU ROKU 2009 A 2010 .....	47
TABULKA 10 - ROZDĚLENÍ ZÁKLADNÍHO SOUBORU HODINOVÉHO POČTU DOTAZŮ .....	48
TABULKA 11 - POČET DOTAZŮ DO JEDNOTLIVÝCH MODULŮ CIS .....	50
TABULKA 12 - POROVNÁNÍ OPTIMALIZOVANÝCH A NEOPTIMALIZOVANÝCH DOTAZŮ .....	55
TABULKA 13 - POROVNÁNÍ ČETNOSTÍ VÝSKYTU ODPOVĚDÍ A DÉLKY ZPRACOVÁNÍ DOTAZU .....	63
TABULKA 14 - DESETI NEJPOUŽÍVANĚJŠÍCH OBRAZOVEK V CIS A JEJICH ČETNOST .....	78
TABULKA 15 - ČETNOST DOTAZŮ DLE DRUHU .....	79
TABULKA 16 -POČET DOTAZŮ DLE DÉLKY ZPRACOVÁNÍ .....	81
TABULKA 17 - POROVNÁNÍ DLE ROZDĚLENÍ BALANCERSERVERU. ....	84

# 1 Úvod

Začátky zavádění výpočetní techniky u cizinecké policie jsou ve znamení nadšenců, kteří na základě svých vědomostí implementovali počítač do života pár jedinců. Jednalo se především o řídicí pracovníky, kteří ve valné většině používali výpočetní techniku jako luxusní psací stroj. Postupem času byly kladeny vyšší nároky na rychlost vyhledávání v bezpočtu evidencí. Jen cizinecká policie udržovala evidenci cizinců trvale a dlouhodobě žijících na území České republiky, evidenci žadatelů o azyl, evidenci cizinců umístěných ve sběrných táborech, evidenci cizinců, kterým byl udělen zákaz pobytu, a mnoho dalších podružných evidencí. Nekonečné řady regálů, páternosterů a registračních skříní zaplňovaly čím dál tím větší prostory státních budov.

Koncem devadesátých let minulého století začal první systematický tlak na plošné využití výpočetní techniky především se zaměřením na evidenční mechanismy s hlavním cílem centralizovat informace a urychlit jejich dostupnost. Pracovníci útvarů informační podpory vyvíjeli programy na evidenci údajů dle jednotlivých problematik. Na svět se dostalo tímto způsobem několik desítek evidenčních programů, které v uvedené době uspokojivě plnily očekávání.

Po roce 2002 byl vyvíjen stále větší tlak na „profesionalitu“ nasazování výpočetní techniky a implementaci programového vybavení. Pod pojmem profesionalita si vedení resortu představovalo svěření tohoto úkolu do rukou odborníků soukromých firem s vidinou úspory finančních nákladů. Přišlo období, kdy se začal prosazovat OUTSOURCING<sup>1</sup>, jako jedna z možností, jak provozovat informační technologii. Tak se to stalo i u cizinecké policie, kdy v roce 2004 implementovala soukromá firma Cizinecký informační systém (dále jen CIS), který byl plně uveden do provozu v roce 2005.

Zadáním výroby evidenčních programů nerezortním subjektům se Ministerstvo vnitra a potažmo Policie České republiky zbavila odpovědnosti za případnou nefunkčnost

---

<sup>1</sup> BRUCKNER, T. - VOŘÍŠEK, J. *Outsourcing informačních systémů*. EKOPRESS, 1998, ISBN 80-86119-07-6.

provozu evidenčních programů a vyhnula se situaci, kdy rezortní programátoři z jakéhokoliv důvodu opustí zaměstnání a nezanechají následovníky. Samotná praxe ovšem prokázala, že finanční prostředky touto cestou ušetřit nelze. Nestabilita legislativy na úseku pobytu cizinců v České republice stále iniciuje realizaci nových a nových změn v programovém vybavení a komunikaci s okolními státy Evropské unie<sup>2</sup>. Neustálá úprava programového vybavení jednotlivých modulů, výstupů a komunikačního rozhraní je stále zvyšující se finanční zátěž tohoto rezortu. Velikou roli zde hraje i nesourodost programového vybavení okolních států, které jsou vázány k vzájemné komunikaci jednotlivými rezolucemi Schengenské úmluvy<sup>3</sup>.

Jednotlivé moduly Cizineckého informačního systému disponují dostatečnou šířkou a kvalitou nástrojů k výběru relevantních dat týkajících se všech okruhů cizinecké problematiky. Dotazovací techniky a výstupní mechanismy jsou na uspokojivé výši i pro potřeby analytických porovnaní jakéhokoliv pohybu migrace nebo signalizací k negativním jevům ze strany cizinců. Relevantní data o pohybu uživatelů v CIS, jejich vytiženost, rychlost zpracování dotazu, využívání neoptimalizovaných dotazů a jejich vliv na chod systému nejsou k dispozici. Při specifikaci zadání programu nebyly tyto údaje preferovány, jelikož v předešlém rezortním zpracování software byly autorovi k dispozici standardně s přístupem ke zdrojovému kódu.

Autoři a prodejci zakázkových evidenčních programů velice neradi přiznávají programovou chybu, nekorektní chování systému, časové prodlevy nad stanovený limit a pozdní reakci při reklamaci funkčnosti. Vzájemná jednání mezi autorem informačního systému a zadavatelem jsou v některých případech velice komplikovaná a často jejich úspěch závisí právě na relevantních a fundovaných informacích inkriminovaného systému. Především vlastnosti systému, jako je například nestabilita, se obtížně dokazují bez jakékoliv vyčíslitelné měrné jednotky.

---

<sup>2</sup> EU. *Smlouva o Evropské unii a smlouvy o fungování Evropské unie v konsolidovaném znění ze dne 15.dubna 2008*

<sup>3</sup> *Schengenská prováděcí úmluva v konsolidovaném znění ze dne 5. dubna 2010*

Jak získat ze strany zadavatele vlastní kanál k zjištění relevantních informací o provozu, zatíženosti, funkčnosti, dodržení stanovených parametrů a využitelnosti informačního systému zakoupeného a provozovaného způsobem Outsourcing, je právě cílem této diplomové práce. Prostředek, který je tím relevantním kanálem k datům, je bezpečnostní logovací soubor Cizineckého informačního systému. Rozbor dat z logovacího souboru za účelem získání relevantních informací, jejich zpracování, vytvoření výstupů vypovídajících o činnosti systému a jejich zavedení do rutinní praxe je nosným pilířem celé práce.

## 2 Cíl práce a metodika

Cílem diplomové práce je nalezení kanálu relevantních dat vypovídajících o funkčnosti provozu informačních systémů, které jsou provozovány Outsourcingovou metodou. Dílčím cílem je vytvoření konkrétních funkčních postupů zpracování těchto dat, jejichž výstupem je seznam informací nebo grafické znázornění. Jak již bylo řečeno v úvodu práce, hlavním nosným pilířem je bezpečnostní logovací soubor, který je nositelem rozsáhlého množství použitelných dat.

Ve stávajícím stavu bezpečnostní log soubory využívá jen gestor CIS odpovídající za shromažďování, ukládání a výdej osobních a citlivých dat. Současný účel bezpečnostních log souborů je tedy monitorování jednotlivých uživatelů pohybujících se v informačním systému. Samotné monitorování uživatelů je základní podmínkou provozu informačního systému s osobními a citlivými údaji, kterou upravuje zákon 101/2000 Sb.<sup>4</sup> o ochraně osobních údajů a o změně některých zákonů ze dne 4. dubna 2000, kde HLAVA II v § 5 přesně vymezuje práva a povinnosti správce.



Obrázek 1 - znázornění navrhovaného řešení

<sup>4</sup> URL: < <http://aplikace.mvcr.cz/archiv2008/sbirka/2000/sb032-00.pdf> > [cit. 2011-12-12]



V levé části obrázku číslo 1 je znázorněn současný stav, kdy přístup k datům přináležejí jen jednomu pracovníkovi odpovědnému za bezpečnou správu dat v CIS. K dispozici má nástroj, ve kterém lze vyhledat jednotlivé dotazy uživatelů, dotazy k určité osobě a s tímto související informace. Vedoucí pracovníci, analytici ani pracovníci podpory informačních systémů mateřského útvaru k datům pro svoje potřeby přístup nesdílí.

Shromažďované údaje o pohybu uživatelů v informačním systému disponují obrovským množstvím dat, které lze bez dalších nákladných prostředků lehkým způsobem využít k zcela jiným účelům, než je jejich původní cíl. Získat lze především velice přesné statistické údaje o rozložení, zatíženosti, využitelnosti a chybách Cizineckého informačního systému. Využitelnost těchto údajů je především v řídicích, analytických a diagnostických směrech provozu útvarů služby cizinecké policie na celém území České republiky, jak je znázorněno v pravé části obrázku číslo 1.

Prvá část diplomové práce je zaměřena na obecné pojmy, metody a předpoklady využití při dalších postupech, výpočtech a měření. Vzhledem k tomu, že měsíční objem log souborů se pohybuje kolem 1,5 až 2,5 GB dat uložených v prostém textovém souboru, je nevyhnutelné se věnovat i časové náročnosti jednotlivých procesů. Jejich optimalizace je součástí všech následných postupů.

V druhé části bude provedeno vyhodnocení jednotlivých dat log souborů a začlenění do struktury podle vypovídající hodnoty. Data je nutno zkonvertovat z prostého textového souboru do struktury, ve které lze realizovat jednotlivé dotazy k zjišťovaným aspektům. Zpřístupnění dat pro účely diplomové práce gestor podmiňuje odfiltrováním jakýchkoliv dat o konkrétní osobě cizince. Odfiltrování dat bude prováděno právě při transformaci dat z textové podoby do struktury relační databáze.

Třetí část práce již řeší konkrétní podobu dotazů do relační databáze za účelem získání vypovídajících dat o jednotlivých zjišťovaných údajích. Hlavní důraz je soustředěn na optimalizovaný a funkční dotaz do relační databáze s následným vyhodnocením relevantnosti zpracovaných informací.

Poslední část diplomové práce vyhodnocuje využitelnost jednotlivých výstupů, začlenění do jednotlivých řídicích, analytických a diagnostických segmentů s možností implementace do každodenní činnosti útvarů služby cizinecké policie.

## 2.1 Metodika

K dosažení hlavního cíle diplomové práce, jímž je získání kanálu k informacím o korektní funkčnosti informačního systému, který spravuje jiný subjekt, bylo použito mnoho metod, postupů a praktik. Krátkým rozbohem těchto nástrojů autor představí zdroje, způsob použití a důvod, proč byly použity uvedené prostředky.

### 2.1.1 Práce s relační databází

Logovací data Cizinecký informační systém zapisuje do prostého textového souboru, který je uložen odděleně od ostatních dat v bezpečném prostoru úložiště dat. Vzhledem k rozsahu dat, jejich bezpečnosti a způsoby vyhledávání k dalšímu zpracování, byla použita transformace do formy relační databáze. Vědomosti o relační databázi získal autor především studiem na České zemědělské univerzitě v Praze v předmětu Databázové systémy<sup>5</sup>, vlastním studiem odborné literatury<sup>6 7</sup> a vyhledáním informací na webových stránkách relevantních zdrojů distributorů konkrétních aplikací<sup>8</sup>.

Přímé využití v diplomové práci je v podobě vytvoření nové struktury tabulky v relační databázi, naplnění daty a realizace výběrů potřebných k dalším výpočtům k získání relevantní informace. Popis transformace z textové podoby dat do formy relační databáze a její následné využití je předmětem další kapitoly.

---

<sup>5</sup> VOSTROVSKÝ, Václav. *Vytváření databází v Oracle*. Česká zemědělská univerzita v Praze. 2008. ISBN 978-80-213-1191-6.

<sup>6</sup> TAYLOR, A. G. *SQL for dummies. 5th edition*. Indianapolis Willey Publishing. ISBN 07645-4075-0

<sup>7</sup> HUMPHRIES, M., HAWKINS, M. a kol. *Data warehousing - Principy a praxe*. Computer Press 2001. ISBN 80-722-6560-1

<sup>8</sup> URL: <[http://technet.microsoft.com/en-us/library/dd349798\(v=WS.10\).aspx](http://technet.microsoft.com/en-us/library/dd349798(v=WS.10).aspx)> [cit. 2011-12-12]

## 2.1.2 Využití statistických metod

Předmětem této práce není statistický rozbor získaných dat, ale jejich možné využití v praxi vedení cizinecké policie a v dohledu nad informačním systémem. Statistické metody jsou při práci využity jen jako prostředek k vytvoření a seskupení dat do podoby, za které se z nich stávají relevantní informace. Z tohoto důvodu autor v postupech využívá pouze základní statistické metody. Cizinecký informační systém vyprodukuje za jeden rok kolem 30 milionů logů o dotazech a kolem 350 milionů logů v podobě konkrétní odpovědi. Data je z tohoto důvodu nutno seskupovat, průměrovat a k možnému vzájemnému porovnání vypočítávat jejich relativní četnost, váhy, nebo procentuální zastoupení. Ve výstupech je poté využito její absolutní nebo kumulativní podoby.

Jeden z bezpečnostních hardwarových prvků je tzv. BalancerSystem, který plní funkci rovnoměrného zatížení obou aplikačních serverů a v případě výpadku jednoho z nich, přeměrování veškeré činnosti na funkční server. K porovnání absolutní četnosti dotazů za určitá časová období je použit korelační koeficient, jehož hodnota v několika situacích vypovídá o závislosti obou serverů. Tyto metody jsou použity hlavně při diagnostických výstupech.

Při rozboru vytíženosti CIS v průběhu časových jednotek (den, týden, měsíc, rok) byla využita metoda kvantilového rozdělení dat na deset částí (decil) a porovnání rozložení souměrnosti číselné řady.

Vědomosti z uvedené problematiky získal autor především studiem na ČZU v předmětech Statistické metody I<sup>9</sup> a Statistické metody II<sup>10</sup>. Konkrétní metody jednotlivých funkcí v prostředí tabulkového procesoru MS Excel zjišťoval autor na webových stránkách podpory výrobce<sup>11</sup>.

---

<sup>9</sup> SVATOŠOVÁ, Libuše – KÁBA, Bohumil. *Statistické metody I*. Česká zemědělská univerzita v Praze. 2009. ISBN 978-80-213-1688-0.

<sup>10</sup> SVATOŠOVÁ, Libuše – KÁBA, Bohumil. *Statistické metody II*. Česká zemědělská univerzita v Praze. 2008. ISBN 978-80-213-1672-0.

<sup>11</sup> URL: < <http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx> > [cit. 2011-12-12]

### 2.1.3 Automatizace některých procesů

V průběhu provádění transformace dat z podoby textové do podoby tabulek relační databáze a v mnoha dalších postupech je využito skriptovacího jazyku k automatizaci této činnosti a k přípravě rutinního provozu. Celá státní správa a potažmo i Policie České republiky staví provoz výpočetní techniky na platformě Microsoft. V návaznosti na tuto skutečnost byl použit skriptovací jazyk Visual Basic Script, který je plně funkční a odladěný v prostředí operačního systému MS Windows 2003 Server. Základní obecné vědomosti získal autor při studiu na ČZU v rámci bakalářského a magisterského programu ze spektra všech odborných předmětů a vlastním studiem odborné literatury<sup>12 13</sup>. Konkrétní syntaxi jednotlivých příkazů a jejich atributů poté na jednotlivých webových stránkách podpory výrobce nebo odborných komunit.

### 2.1.4 Práce s výstupy

Bezpečnostní logovací soubor je řada oddělených dat, která při pohledu laika nedávají žádný smysl. K využití těchto dat na relevantní informace použitelné k řídicím, analytickým a diagnostickým účelům je nutné vytvořit intuitivní výstupy. Mezi takovéto výstupy patří především grafické znázornění jednotlivých vypočítaných hodnot, nebo jejich zpracování do tabulky.

V diplomové práci byly použity především nástroje tabulkového editoru MS EXCEL verze 2010. Při konkrétním využití jednotlivých funkcí a nástrojů, byla použita nápověda programu a on-line podpora výrobce.

---

<sup>12</sup> MULLER, Scott. *Osobní počítač*. Computer press. 2001. ISBN 80-7226-470-2.

<sup>13</sup> HOMER, Alexandr, BRIAN, Francis a kol. *Aktive Server Pages 3.0.*, Computer press. 2000. ISBN 80-8609-747-1.

## 2.1.5 Práce s textovým dokumentem

Textový dokument byl vytvářen v prostředí textového editoru MS WORD 2010 v českém jazyce. K formátování diplomové práce byly použity nejrůznější pomůcky a doporučení získané na oficiálních webových portálech České zemědělské univerzity v Praze (student.czu.cz, moodle.czu.cz). Další zdroje byly získány v komentované normě ČSN ISO 690 a ČSN ISO 690-2.<sup>14 15</sup>

---

<sup>14</sup> BOLDYŠ, Petr. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 1 – Citace: metodika a obecná pravidla*. Verze 3.3. c1999-2004, poslední aktualizace 11.11.2004. URL: <<http://WWW.boldis.cz/citace/citace1.pdf>>.

<sup>15</sup> BOLDYŠ, Petr. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 2 – Modely a příklady citací u jednotlivých typů dokumentů*. Verze 3.0 (2004), poslední aktualizace 11.11.2004. URL: <<http://WWW.boldis.cz/citace/citace2.pdf>>.

### 3 Přehled řešené problematiky

Řešená problematika obsahuje několik pojmů, které je nutno objasnit před samotným využitím. Jedná se například o pojem a metodu outsourcing, který je stěžejním důvodem, proč hledat další kanál k informacím o systému, který cizinecká policie vlastní.

#### 3.1 Outsourcing, historie a uplatnění u cizinecké policie

*Informační systémy a informační technologie (IS/IT) se v 90. letech dvacátého století staly jedním z nejvýznamnějších zdrojů hospodářských organizací<sup>16</sup>. Vyspělé hospodářské organizace vynakládají v posledních letech na IS/IT vysoké finanční částky (až 20% z celkového objemu investic). S růstem významu IS/IT roste i význam kvality řízení IS/IT hospodářské organizace. Návrh architektury informačního systému, realizace jednotlivých stavebních kamenů této architektury formou informatických projektů, údržba a provoz jednotlivých aplikací jsou činnosti, které jsou velmi náročné jak na kvalifikaci pracovníků, tak na celkový objem lidských a finančních zdrojů. Pro mnohé hospodářské organizace*



Obrázek 2 – znázornění rozdělení odpovědnosti za provoz CIS

<sup>16</sup> BRUCKNER, T. - VOŘÍŠEK, J.: *Outsourcing informačních systémů*. s.10. EKOPRESS, 1998, ISBN 80-86119-07-6.

*se ukazuje jako finančně a personálně neúnosné nebo dokonce zcela nemožné, aby si všechny činnosti související s vývojem, provozem a údržbou IS/IT zajišťovaly vlastními silami. Snaží se proto vytěsnit některé činnosti mimo podnik – na externí dodavatele komponent a služeb IS/IT. Jinými slovy snaží se realizovat outsourcing. Jde-li podnik cestou outsourcingu, naráží na řadu netriviálních otázek, jako například: které činnosti je možné vytěsnit, aniž by se podnik stal neúnosně závislým na externích dodavatelích, které činnosti je nejvhodnější vytěsnit s ohledem na celkové ekonomické efekty IS/IT, jak řešit systémovou integraci v podmínkách širokého spektra externích dodavatelů atd.<sup>17</sup>*

Pomocí metody outsourcingu cizinecká policie řešila v letech 2003-2004 vybudování globálního informačního systému pro své útvary. Výsledný produkt pod názvem Cizinecký informační systém byl spuštěn v roce 2005 a od té doby je provozován částečně pracovníky cizinecké policie a soukromou firmou xy.cz<sup>18</sup>. Ve zjednodušené formě lze říci, že uživatelskou část zajišťují pracovníci mateřského útvaru cizinecké policie. Serverovou část, tedy aplikační servery, datové úložiště a chod programového vybavení, zajišťuje firma xy.cz. Za obsahovou část informačního systému, data a nakládání s nimi odpovídá gestor z řad Policie České republiky. Na základě nastavené kompetentnosti a odpovědnosti je dána situace, kdy pracovníci PČR nedisponují přístupem k programovému kódu a jeho vnitřním diagnostickým částem.

## 3.2 Logovací soubor

Dalším základním pojmem v diplomové práci je logovací soubor jako zdroj dat, z kterých se pomocí jednotlivých částí této práce stanou užitečné informace pro řídicí pracovníky jednotlivých útvarů cizinecké policie.

Logovací soubor (EventLog - LogSoubor) nemá svoji obecnou definici formátu. Každá firma si vytváří svůj formát zvlášť, a to i na různé verze svého software. Znamější firmy

---

<sup>17</sup> BRUCKNER, T. - VOŘÍŠEK, J.: *Outsourcing informačních systémů*. s.10. EKOPRESS, 1998, ISBN 80-86119-07-6.

<sup>18</sup> Autor diplomové práce nedostal svolení se zveřejněním názvu autorské soukromé firmy díla CIS, proto bude používat nadále pro potřeby DP smyšlený název **xy.cz**

a organizace formát svého produktu zveřejňují k dalším možným účelům. Mezi takové firmy patří například IBM<sup>19</sup>, Microsoft<sup>20</sup> nebo organizace W3<sup>21</sup>.

Logovací soubor (dále jen LogSoubor) si lze představit jako pomyslný „zápisník“, do kterého si zejména operační systém zaznamenává výskyt nejružnějších událostí.

***Obecně lze rozdělit logy operačního systému na:***

- systémové
- zabezpečení
- aplikační
- instalační

Událostní logy kromě operačního systému využívají i jednotlivé aplikace, programy a utility. Využití logování probíhá všude tam, kde je potřeba zadokumentovat činnosti počítače s možností jejich zpětného vyhodnocení.

***Textová podoba zápisu informací o události***

Při prvním pohledu na zápis jednoho řádku z neznámého LogSouboru lze vyčíst například, že třetí položkou je datum a čas výskytu události, pátá položka by mohla být IP adresa. Další informace, která zbylá data sdělují, jsou však již nesrozumitelné.

```
D|e99f071754a311dfbf0d00212818282e|2010-07-01
00.01.22|cis\xy301493|110.186.35.166|CIS|Q|SS:[PLU],OBR:[Podrobná
lustrace],EV:[EVIC2,TDU,VIZ,UBY],PH:[DL000050P40],JM:[TSEVEGMID],PR:[SU
KHBAATAR],STP:[MNG]|1|||||KONTROLA CIZINCE|CIS|CIS|cisas2|cisas2
```

Takto za sebou jdoucí data nelze řádně vyhodnotit, jelikož není znám jejich formát. Bez dokumentace a popisu jednotlivých dat není možné získat použitelné informace o průběhu činnosti procesů, uživatelů, periférií a podobně.

---

<sup>19</sup>URL: < [http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA\\_info45/en\\_US/HTML/guide/c-logs.html](http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA_info45/en_US/HTML/guide/c-logs.html) > [cit. 2011-12-12]

<sup>20</sup>URL: < [http://technet.microsoft.com/en-us/library/cc786081\(v=ws.10\).aspx](http://technet.microsoft.com/en-us/library/cc786081(v=ws.10).aspx) > [cit. 2011-12-12]

<sup>21</sup> URL: < <http://www.w3.org/Daemon/User/Config/Logging.html#LogFormat> > [cit. 2011-12-12]



### 3.2.1 Struktura logovacího zápisu

Vyhodnocení LogSouborů je jednoduchá záležitost v případě, kdy je znám význam jednotlivých zaznamenaných údajů. Ve většině případů ovšem LogSoubory obsahují celou řadu zkratk, kódů, výpisů z paměti a nejrůznějších dat, které bez podrobného popisu a znalosti zapsaných údajů nelze identifikovat. Je tudíž nutné znát jejich význam a formát zápisu.

LogSoubor má nejčastěji podobu prostého textového formátu. Při výskytu zájmové události v systému se automaticky zapíše do tohoto souboru předem dané informace v podobě dat. Jednotlivá data od sebe v souboru rozlišujeme oddělovačem. Oddělovačem rozumíme jeden či více znaků, které při zápisu do souboru oddělují jednotlivá data. Mezi nejčastěji používané znaky oddělovače patří především středník „;“, „“ a vertikální čára „|“. Pro zápis jedné události se převážně využívá právě jeden řádek textového souboru. Na základě těchto jednoduchých pravidel lze následovně s Log souborem aktivně pracovat a využívat jej jako zdroj základních informací o činnosti počítače, uživatelů, chybách, napadení systému a podobně.

### 3.2.2 Formát zápisu logů

Formát zápisu logů není nikde přesně definován. Obecně lze říci, že se jedná většinou o neformátovaný prostý textový soubor, který v drtivé většině případů používá koncovku „log“. Celkové označení souboru je tedy například „LogSoubor1.log“. Zápis jedné události je obvykle v jednom řádku s oddělovači. Každý zápis události začíná ustanovením data a času události, její název a popis individuální sledované činnosti. Jako příklad si můžeme uvést strukturu Event Logu operačního systému MS Windows verze 7 z kolekce SYSTEM<sup>22</sup>, jehož výpis je uveden v následujícím příkladu.

---

<sup>22</sup> URL: <[http://technet.microsoft.com/en-us/library/dd349798\(v=WS.10\).aspx](http://technet.microsoft.com/en-us/library/dd349798(v=WS.10).aspx)> [cit. 2011-12-12]

```

-   System
      -   Provider
          [ Name]      Microsoft-Windows-Security-SPP
          [ Guid]      {E23B33B0-C8C9-472C-A5F9-F2BDFEA0F156}
          [ EventSourceName]  Software Protection

Platform Service
      -   EventID      902
          [ Qualifiers]  16384
          Version      0
          Level 0
          Task 0
          Opcode      0
          Keywords     0x8000000000000000
      -   TimeCreated
          [ SystemTime]  2011-09-18T07:49:16.000000000Z
          EventRecordID  50750
          Correlation
      -   Execution
          [ ProcessID]   0
          [ ThreadID]   0
          Channel

Application
      Computer      Peking2
      Security

-   EventData 6.1.7600.16385

```

Tučný text označuje vysvětlující název informace, která v LogSouboru není uvedena. Slabě označený text je posléze v tomto pořadí zapsán do log souboru. Jednotlivé údaje jsou od sebe odlišeny oddělovači. Z příkladu je patrné, že uživatel, který nedisponuje popisem položek v log souboru, není schopen interpretovat zápis dat.

*Samotný zápis jednoho řádku v textovém souboru z uvedeného příkladu:*

```
Microsoft-Windows-Security-SPP|{E23B33B0-C8C9-472C-A5F9-  
F2BDFEA0F156}|Software Protection Platform  
Service|902||0|0|0|0|0x8000000000000000|2011-09-  
18T07:49:16.000000000Z|50750|0|0|Application|Peking2|Security|6.1.7600.1  
6385|
```

Hlavním předmětem diplomové práce je rozbor bezpečnostního LogSouboru Cizineckého informačního systému (CIS). Popisem Cizineckého informačního systému a některých jeho částí autor přiblíží funkčnost a důležitost celku a jednotlivých modulů.

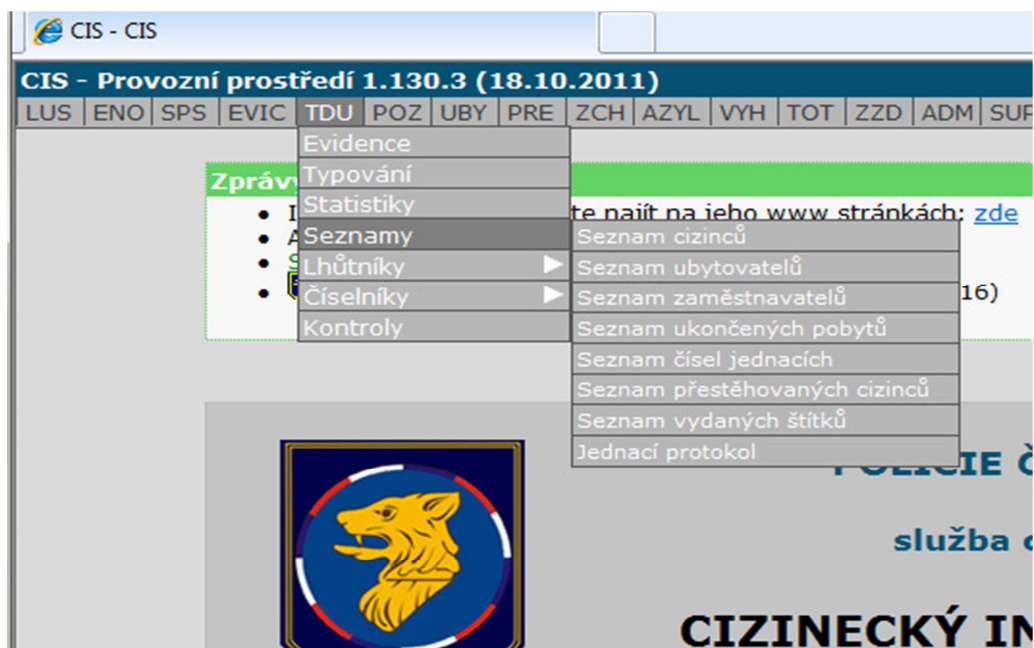
### **3.3 Cizinecký informační systém (CIS)**

Cizinecký informační systém (dále jen CIS) je webová aplikace sdružující všechny evidence pro práci cizinecké policie. Cizinecká policie plní úkoly stanovené především zákonem o pobytu cizinců na území České republiky 326/1999 Sb.<sup>23</sup>, který ji mimo jiné ukládá vést evidenci cizinců a údajů k nim. Položková skladba vedených informací je dosti rozsáhlá a kopíruje požadavky ustanovení uvedeného zákona Hlavy XV § 158 odstavce 1 až 5, který vymezuje vést evidenci v elektronické podobě. Tento normativ je dále rozpracován rozkazem policejního prezidenta.<sup>24</sup> Celá evidence zahrnuje údaje o cizincích v režimu azylu, pobytovém režimu, institut pozvání, ubytování, přestupková řízení, správní vyhoštění, index nežádoucích osob, odcizené cestovní doklady, schengenská víza a podobně. Jednotlivé evidence jsou rozděleny do modulů. Kromě standardních modulů podle problematiky disponuje CIS také modulem „Rychlé lustrace“, který slouží k lustraci na hraničním přechodu. V podmínkách České republiky se jedná o hraniční přechody v areálu mezinárodních letišť. Rozsah Cizineckého informačního systému lze nejlépe dokreslit plným zněním § 158 zákona číslo 326/1999 Sb., o pobytu cizinců na území České republiky, který celý pojednává o povinnostech Policie České republiky na úseku provozu evidence cizinců a je uveden v příloze číslo 5.

---

<sup>23</sup> URL: < <http://aplikace.mvcr.cz/archiv2008/sbirka/1999/sb106-99.pdf> > [cit. 2011-12-12]

<sup>24</sup> Policie České republiky. *Závazný pokyn policejního prezidenta č. 159 ze dne 29. prosince 2004, kterým se upravuje postup při provozování informačního systému Policie České republiky služby cizinecké a pohraniční policie*



Obrázek 3 – ukázka grafického prostředí CIS s rozsáhlou nabídkou vyhledávacích nástrojů

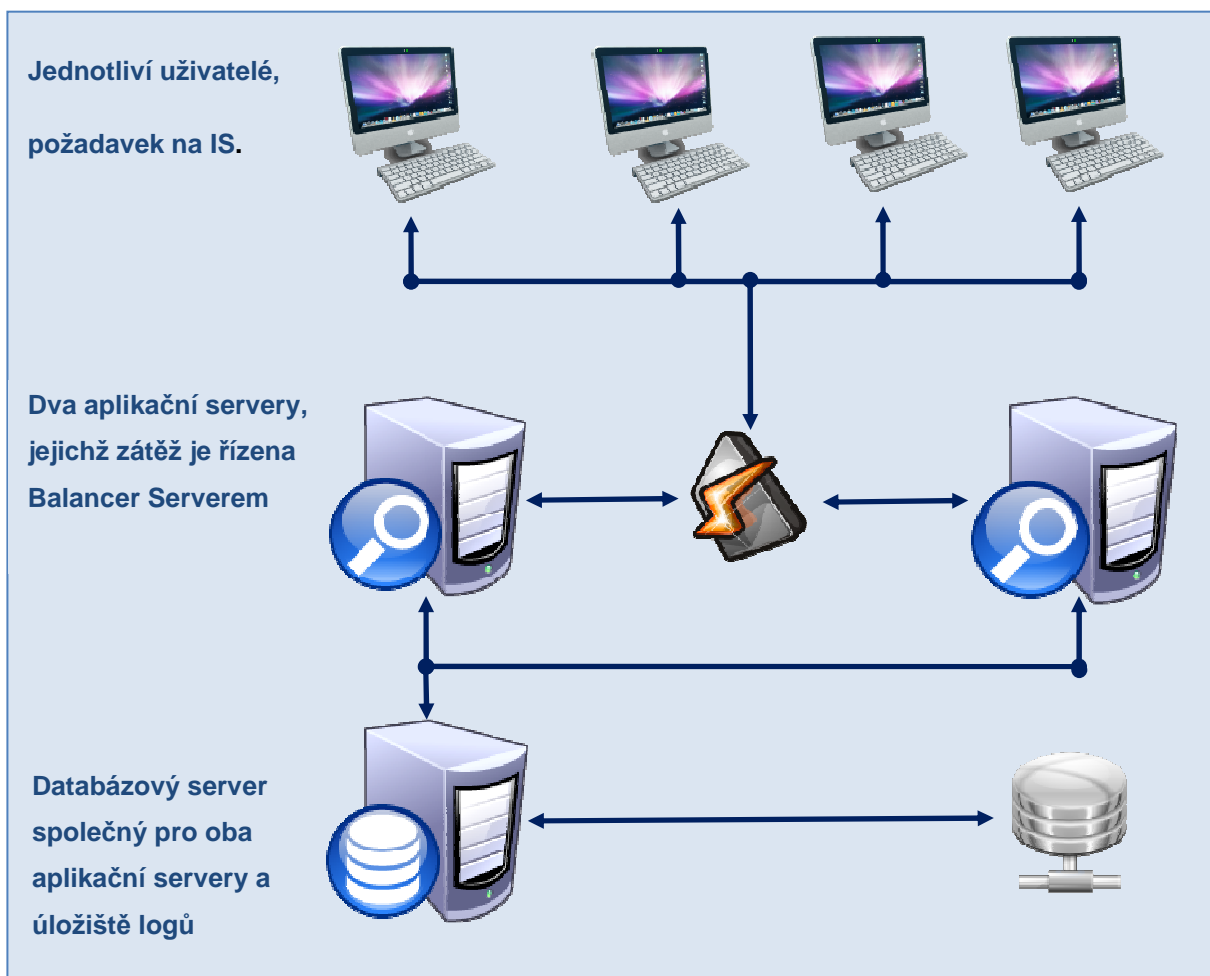
CIS disponuje velkým množstvím nástrojů pro vkládání, editaci, výmaz a výběr dat. Jak je patrné z obrázku číslo 3, který znázorňuje grafické prostředí informačního systému, pro výběr dat je vybaven například funkcí typování, statistikou, seznamy, lhůtníky, kontroly a podobně. Všechny tyto nástroje vyhledávají především data o cizincích a jejich evidovaných údajích.

### 3.3.1 Skladba databáze programu CIS a vnější zdroje

Skladba databáze Cizineckého informačního systému<sup>25</sup> je již plně strukturovaná. Všechny důležité položky systém rozprostře do jednotlivých tabulek s napojením k souvztažným údajům. Základní databáze disponuje více než devadesáti tabulkami, v nichž jsou zaznamenány jednotlivé entity uložených informací, například entita osoby cizince je v standardních evidencích prezentována příjmením, jménem a datem narození, pohlavím, místem narození, státem, údaji o rodičích a potomcích. Pro účely CIS je osoba rozčleněna na entitu identity, kde jsou zaznamenány údaje o datu narození, místu narození

<sup>25</sup>URL: < <http://aplikace.mvcr.cz/archiv2008/sbirka/1999/sb106-99.pdf> > [cit. 2011-12-12]

a několik dalších identifikačních údajů nutných pro propojení s dalšími tabulkami. Takto postavená identita člověka je neměnná a zajišťuje nezaměnitelnost osoby cizince.



Obrázek 4 - znázornění využití Balancer Serveru a ukládání dat v CIS (vytvořil autor)

Všechny ostatní údaje, například jméno, příjmení, pohlaví, rodiče, potomci, jsou považovány za údaje vedlejší, které identitu ovlivní, ale nezmění. Zjednodušeně lze říci, že osoba, která si provede změnu pohlaví, příjmení i jméno je nadále pro informační systém CIS stále jedna identita osoby, jako před změnou. Kromě vlastních údajů využívá CIS i datové zdroje z vnějšího prostředí. Mezi hlavní patří vnitřní evidenční systém Ministerstva zahraničních věcí, který eviduje víza vydaná mimo republiku. Cizinecký informační systém využívá registr staveb katastrálního úřadu ke ztotožnění bydliště, jelikož každá adresa má přiděleno v celostátním systému identifikační číslo. V praxi to znamená, že uživatel nevpisuje do systému adresu, ale je povinen ji vybrat z číselníku. Mezi další datové zdroje, které jsou využívány je například databáze SIS (Shengenský

informační systém), PTROS (pátrání po osobách) a podobně. CIS je rozsáhlá aplikace s centrálním úložištěm databázového systému na platformě UNIX. Komunikaci s webovými prohlížeči a logické procedury zpracovávají aplikační servery, které jsou vybaveny „balancer-systémem“. Systém zajišťuje rozložení síťového přístupu uživatelů na oba aplikační servery rovnoměrně a v případě výpadku jednoho serveru přesměrování na server funkční. Funkčnost celé aplikace je znázorněna na obrázku č. 4.

### 3.3.2 Bezpečnostní LogSoubor v CIS

Na základě ustanovení zákona č. 101/2000 Sb.<sup>26</sup>, o ochraně osobních údajů a o změně některých zákonů ze dne 4. dubna 2000, kde HLAVA II v § 5, přesně vymezuje práva a povinnosti správce programů, které zpracovávají a uchovávají osobní údaje a údaje označeny jako data citlivá.

Jednou ze základních povinností těchto útvarů je evidence přístupu k údajům jednotlivých uživatelů a monitorování s jejich nakládáním. CIS je vybaven nástrojem, který monitoruje veškerou činnost uživatele. O každé činnosti je proveden zápis do zvláštního textového souboru, tzv. bezpečnostního LogSouboru.

### 3.3.3 Rozdělení kompetencí při provozu CISu

Jak již bylo řečeno, CIS není dílem programátorů personálně nebo pravomocně spadajících do kompetence Ministerstva vnitra nebo jeho složky Policie České republiky. CIS byl vyroben a nadále je udržován firmou xy.cz, tzv. „třetí stranou“. Vzhledem k této skutečnosti bylo nutno rozhodnout o konkrétní odpovědnosti a kompetentnosti k jednotlivým částem systému. Zjednodušeným způsobem, který postačí k znalosti základních zásad a principů, lze říci, že za chod Cizineckého informačního systému po stránce hardwarové a softwarové odpovídá firma, která CIS vytvořila. (obrázek č. 5)

---

<sup>26</sup> URL: <<http://aplikace.mvcr.cz/archiv2008/sbirka/2000/sb032-00.pdf>> [cit. 2011-12-12]

Ministerstvo vnitra, respektive Policie České republiky, odpovídá za data, která jsou v systému shromažďována, ukládána a dále distribuována.



Obrázek 5 – znázornění rozdělení kompetencí při provozu CIS

Z toho plyne, že uživatelé a obslužný personál nemá přístup ke zdrojovým kódům samotného programu. Pracovníci firmy xy.cz nemají přístup k datům, týkajících se údajů o konkrétní osobě cizince. Z těchto skutečností je patrné, že uživatelé a obslužný personál nedisponuje přístupem k aplikačním a systémovým LogSouborům. Ve vlastním zájmu správce dat zase pracovníci civilní firmy nedisponují přístupem k bezpečnostnímu LogSouboru, který v sobě nese a shromažďuje osobní údaje.

Na základě těchto skutečností byl hledán kanál k datům, která by po zpracování podávala relevantní informace o některých nedostupných údajích z Cizineckého informačního systému. Tímto kanálem je právě bezpečnostní LogSoubor CIS.

## 3.4 Struktura bezpečnostního LogSouboru CIS

Struktura bezpečnostního LogSouboru CIS se skládá ze tří základních parametrů. Údaje parametru **dotazu**, **odpovědi** a **přílohy**. Příloha obsahuje identifikační údaje osob, které jsou aplikačním serverem vráceny a zobrazeny u uživatele. Jako oddělovač mezi daty je použit znak „|“.

*Příklad jednoho celku dotazu, odpovědi a přílohy:*

**D** | aaad06354a311dfbf0d00212818282e | 2010-07-01  
00.00.09 | cis\rx301111 | 101.186.3.1 | CIS | Q | SS: [PLU], OBR: [Podrobná  
lustrace], EV: [EVIC2, TDU, VIZ, UBY], PH: [DL000005AAAA], DN: [25.05.1980], PR: [CHIMEG] | 1 | | | | KONTROLA CIZINCE | CIS | CIS | cisas2 | cisas2

**O** | aaaad06354a311dfbf0d00212818282e | 1 | 2010-07-01 00.00.10 | 3

**P** | aaaad06354a311dfbf0d00212818282e | LT | 11119796 | CHIMEG  
PUNTS | 19800525 | | M | MAG

**P** | aaaad06354a311dfbf0d00212818282e | LV | 11113471 | CHIMEG  
PUNTS | 19800525 | | M | MAG

**P** | aaaad06354a311dfbf0d00212818282e | LV | 11110718 | CHIMEG  
PUNTS | 19800525 | | M | MAG

### 3.4.1 Struktura logu dotazu („D“):

Věta dotazu začíná identifikátorem „D“. V dotazu jsou zaznamenány všechny informace o zadání dotazu uživatelem do informačního systému. Zaznamenán je především čas, kdy byl dotaz zaslán ke zpracování. Tento údaj je velice důležitý k zjišťování délky zpracování příkazu aplikačním serverem, jelikož další časová stopa je uvedena ve struktuře odpovědi z aplikačního serveru. Odečtením těchto dvou údajů je získána doba zpracování požadavku uživatele. Jedná se o jednu z nejrelevantnějších informací. Datová struktura je uvedena v tabulce č. 1.



Tabulka 1 - struktura dotazu "D"

Por.	Význam položky	Možnosti
1	identifikátor	"D" - dotaz, "O"- odpověď, "P"-přílohy
2	jedinečný klíč dotazu	Spojuje údaje dotazu, odpovědi a přílohy
3	datum a čas dotazu	
4	identifikace uživatele	
5	IP adresa PC	
6	identifikace IS	
7	znak pro úkon uživatele	"Q" lustrace, "I" zápis, "D"výmaz, "U"změna
8	údaje dotazu ve formuláři	modul, obrazovka, druh evidence, výpis údajů z formuláře
9	pokus	v případě neúspěchu pořadí dalšího pokusu
10	bezvýznamná data	
11	bezvýznamná data	
12	bezvýznamná data	
13	bezvýznamná data	
14	bezvýznamná data	
15	důvod kontroly	důvod lustrace cizince
16	označení IS	
17	označení IS	
18	označení serveru balanceru	cisas1, cisas2
19	označení serveru balanceru	cisas1, cisas2

### 3.4.2 Struktura logu odpovědi („O“):

Struktura logu odpovědi zaznamenává data o času vrácení údajů na dotaz od serveru k uživateli, jeho zdárné zpracování a počet odpovědí. Význam jednotlivých údajů logu odpovědi popisuje autor v tabulce číslo 2.

Tabulka 2 - struktura logu odpovědi "O"

Por.	Význam položky	Možnosti
1	identifikátor věty	"O"- odpověď
2	jedinečný klíč dotazu	Spojuje údaje dotazu, odpovědi a přílohy
3	druh odpovědi	standard ne chybová
4	datum a čas odpovědi	datum a čas vrácení odpovědi po zpracování dotazu serverem
5	počet odpovědí	

### 3.4.3 Struktura logu příloh („P“):

Log příloh uchovává data o vrácených cizincích s odpovědi a disponuje především údaji o ID cizince, příjmení a jménu, evidenci, data narození a státní příslušnost. Význam jednotlivých údajů logu odpovědi jsou popsány v tabulce číslo 3.

Tabulka 3 - struktura logu příloh "P"

Por.	Význam položky	Možnosti
1	identifikátor	"P" - příloha odpovědi
2	jedinečný klíč dotazu	Spojuje údaje dotazu, odpovědi a přílohy
3	identifikace modulu IS	LT, LV a podobně
4	ID osoby	
5	pomocné parametry	
6	jméno a příjmení osoby	
7	data narození	formát RRRRMMDD
8	rodné příjmení	rodné příjmení v identitě
9	pohlaví	M -muž, Z-žena, X-nezjištěno
10	bezvýznamná data	
11	státní příslušnost	dle mezinárodního číselníku ve formátu SSS

### 3.4.4 Objem dat bezpečnostních LogSouborů CIS

Pro další výpočty a návrh relační databáze byl zjištěn objem dat, která uchovávají LogSoubory za dva poslední roky. Skriptem pro jednoduchý součet jednotlivých logovacích vět v souborech, především D-dotazů, O-odpovědí a P-příloh, bylo zjištěno, že za dva roky vyprodukuje CIS celkem 381 milionů záznamů, což je kolem 15 milionů za měsíc. V dalších výpočtech je nutné předpokládat práci s větším objemem dat a k tomu přizpůsobit relační databázi. V následující tabulce jsou konkrétní součty vět za jednotlivé roky s výpočtem průměru na jeden měsíc.

Tabulka 4 - objem dat v bezpečnostním LogSouboru CIS

Objem dat LS CIS	pocetD	pocetO	pocetP	pocetJ	pocetO+D+P+J	O+D+P+J/12
<b>Součty celkem:</b>	51911026	51911021	277000931	5393	380828371	
<b>Součty za rok 2009</b>	24996307	24996305	126944910	3136	176940658	14745055
<b>Součty za rok 2010</b>	26914719	26914716	150056021	2257	203887713	16990643

## 3.5 Struktura relační databáze a výběr dat z LogSouboru

Jedním z hlavních úkolů diplomové práce je výběr relevantních dat z LogSouborů CIS pro další výpočty. Vybraná data je nutno posléze přetransformovat do struktury relační databáze k rychlejším, přehlednějším a spolehlivějším výpočtům. Záměrem je vytvoření jedné jediné tabulky, která bude obsahovat všechna relevantní data k další práci. Po několika výběrech a pokusech s daty vznikla tabulka relační databáze, která je znázorněna v podobě SQL příkazu pro tvorbu relační tabulky.

```
CREATE TABLE [dbo].[tLog] (  
    [nLogId] [numeric](18, 0) IDENTITY (1, 1) NOT NULL ,  
    [dLogDateTimed] [datetime] NULL ,  
    [nLogDelkaZprac] [numeric](18, 0) NULL ,  
    [cLogUserNameD] [char] (8) COLLATE Czech_CI_AS NULL ,  
    [cLogIpD] [char] (15) COLLATE Czech_CI_AS NULL ,  
    [cLogDruhD] [char] (1) COLLATE Czech_CI_AS NULL ,  
    [tLogDotaz] [text] COLLATE Czech_CI_AS NULL ,  
    [cLogMod] [char] (10) COLLATE Czech_CI_AS NULL ,  
    [cLogObr] [char] (100) COLLATE Czech_CI_AS NULL ,  
    [nLogPocet] [numeric](10, 0) NULL ,  
    [bLogDuvod] [bit] NULL ,  
    [bLogServer] [bit] NULL ,  
    [rLogDen] [char] (1) COLLATE Czech_CI_AS NULL)  
ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]
```

### 3.5.1 Jednotlivé položky relační tabulky

Část položek je přímo kopírována z údajů LogSouboru do struktury relační databáze. Některá data jsou však při transformaci vypočítávána. Konkrétní podobu dat a jejich výpočet je uveden v stručném popisu jednotlivých položek nově vzniklé tabulky.

*Stručný popis jednotlivých položek vzniklé relační tabulky:*

**nLogId** – jedinečné označení řádku v tabulce, klíč tabulky

**dLogDateTimeD** – zkopírovaný datum a čas počátku zpracování dotazu ze str. logu „D“

**nLogDelkaZprac** – jedná se o rozdíl času odpovědi z str. logu „O“ a počátku času zpracování dotazu ze str. logu „D“, tudíž délka zpracování dotazu

**cLogUserNameD** – zkopírovaná identifikace uživatele dotazu str. logu „D“

**cLogIpD** – zkopírovaná identifikace počítače z str. logu „D“

**cLogDruhD** – zkopírovaný kód druhu dotazu z str. logu „D“ (Q,I,U)

**tLogDotaz** – zkopírovaný celý text dotazu uživatele z str. logu „D“, pol. 8

**cLogMod** – název modulu, který je dotazován, select z str. logu „D“, pol. 8

**cLogObr** – název obrazovky z které je dotaz, select z str. logu „D“, pol. 8

**nLogPocet** – počet odpovědí na dotaz, copy z str. logu „O“, položka 5

**bLogDuvod** – (0,1) zapsán důvod lustrace, nezapsán důvod lustrace

**bLogServer** – (0,1) server, který zpracovává dotaz, 0-první server, 1-druhý server

**rLogDen** – číselné označení dne v týdnu (pondělí – 1, neděle - 7)

Konkrétní výpočty a výběry jsou patrné přímo ze skriptu, který vybírá a převádí data z textové podoby do relační databáze.

### **3.6 Transformace dat z LogSouboru do relační databáze**

Pro další výpočty a práci s logy je provedena transformace dat z textové podoby do struktury relační databáze. Při transformaci jsou především vybírána relevantní data, provedena jejich kontrola a provedena příprava k exportu do databáze. Transformačním

nástrojem je skriptovací jazyk Visual Basic Script (dále jen VBScript)<sup>27</sup>. Kód skriptu, který je opatřen popisem funkcionality jednotlivých částí, uvádí autor v následující části.

### ***Kód skriptu transformace dat***

```
Option Explicit
ON ERROR RESUME NEXT
Dim a,b
a=0
b=0
Dim sPathFolder          ' domovsky adresar
sPathFolder=WScript.ScriptFullName
sPathFolder=Mid(sPathFolder,1,Len(WScript.ScriptFullName)-
Len(WScript.ScriptName))
Dim dokumentTxt          ' text s cim porovnavame (vetsi)
dokumentTxt = sPathFolder + "MesicVstup.txt" ' ----->> ZADAT
Dim fso ' files system object
Dim objSoNew ' objekt otevreného souboru tabulky
Dim objSoDokument ' objekt vytvoreneho souboru noveho dokumentu
Dim objConn, objRst
' vytvori objekt FileSystemObject
Set fso = CreateObject("Scripting.FileSystemObject")
' otevre dokument tabulku pro cteni
Set objSoDokument= fso.OpenTextFile(dokumentTxt, 1, False)
' otevre New tabulku pro psani
DIM text, text2, SpText, i, vystup
text=""
text2=""
DIM vdLogDateTimed, vnLogDelkaZprac, vcLogUserNameD, vcLogIpD,
vcLogDruhD
DIM vtLogDotaz, vcLogMod, vcLogObr, vnLogPocet, vbLogDuvod, vbLogServer,
vrLogDen
' connection string pro pstup k databzi
const objConnStr = "Provider=SQLOLEDB.1; Data Source=mvcr-75e18b80e5;
Initial Catalog=simi; User Id=sa; Password=a"
' Zapis do DBF
set objConn = CreateObject("ADODB.Connection")
objConn.Open(objConnStr)
while Not objSoDokument.AtEndOfStream
    text = TRIM(objSoDokument.ReadLine)
    a=a+1
    IF LEFT(text,2)="D|" THEN
        text2 = TRIM(objSoDokument.ReadLine)
        IF Mid(text,3,32) = Mid(text2,3,32) AND LEFT(text2,
2)="O|" THEN
            text = text2 & "|" & text
            SpText = Split(text, "|", -1,1)
            b=b+1
        ' zadani jednotlivych promenych a priprava SELECT
vdLogDateTimed = Replace(SpText(7),".",":")
vcLogUserNameD = Right(SpText(8), 8)
vcLogIpD = Left(SpText(9),15)
vcLogDruhD = Left(SpText(11), 1)
vtLogDotaz = REPLACE(Replace(SpText(12),"@","s"), "","_-")
vcLogMod = Prohledat(vtLogDotaz, "SS:[" , "]"")
vcLogObr = Prohledat(vtLogDotaz, "OBR:[" , "]"")
vnLogPocet = SpText(4)
IF SpText(19) <> "" THEN
    vbLogDuvod = 1
ELSE
    vbLogDuvod = 0
```

<sup>27</sup> URL: <[http://msdn.microsoft.com/en-us/library/3ca8tfek\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/3ca8tfek(v=vs.85).aspx)> [cit. 2011-12-12]

```

END IF
IF SpText(22) = "titan" OR SpText(22) = "cisas2" THEN
    vbLogServer = 1
ELSE
    vbLogServer = 0
END IF
vrLogDen = weekDay(CDate(Left(SpText(7), 10)), 0)
vnLogDelkaZprac=9999
    vnLogDelkaZprac = CInt((CDate(Replace(SpText(3), ".", ":")) -
    CDate(Replace(SpText(7), ".", ":")))*(3600*24))
    set objRst = objConn.Execute("INSERT INTO tLog(dLogDateTimed,
    nLogDelkaZprac, cLogUserNameD, cLogIpD, cLogDruhD, tLogDotaz,
    cLogMod, cLogObr, nLogPocet, bLogServer, bLogDuvod, rLogDen)
    VALUES('"&vdLogDateTimed&"', '"&vnLogDelkaZprac&"', '"&vcLogUserNameD
    &"', '"&vcLogIpD&"', '"&vcLogDruhD&"', '"&vtLogDotaz&"',
    '"&vcLogMod&"', '"&vcLogObr&"', '"&vnLogPocet&"',
    '"&vbLogServer&"', '"&vbLogDuvod&"', '"&vrLogDen&"')")
set objRst = nothing
    END IF
    END IF
' konec cyklu tabulky objSoDokument
wend
' uzavreni objektu
Set objSoNew = Nothing
Set objSoDokument = Nothing
Set fso = Nothing
set objConn = Nothing
WScript.echo "Celkem: " & a & " Zpracovano: " & b & "Cas: " & Time()
WScript.quit
Function Prohledat(kde, zac, kon)
    ' funkce Prohledat() prohleda retezec "kde" a bude hledat předponu
    "zac" a ukončení "kon"
    ' příklad (retezec --> .....COM:[nejaky
    text]..... ) zac="COM:[
    --> vysledek je "nejaky retezec" kon="]"
    Dim start2
    IF InStr(kde, zac) > 0 THEN
        start2=InStr(InStr(kde, zac), kde, kon)
        Prohledat=Mid(kde, InStr(kde, zac) + Len(zac), start2-
        (InStr(kde, zac)+Len(zac)))
    ELSE
        Prohledat = "nenalezeno"
    END IF
END Function

```

### 3.6.1 Časová náročnost transformace dat

LogSoubory vytváří každý ze dvou aplikačních serverů. Měsíční objem dat z jednoho serveru je kolem 8 až 15 milionů řádků v prostém textovém formátu. Transformace jednoho LogSouboru podle měření počátku a konce procedury trvá v průměru kolem 11 minut. Měsíčně je potřeba sehrát dva LogSoubory. Časová náročnost sehrání dat je tedy do půl hodiny měsíčně. Transformace dat byla realizována na standardním serveru

střední třídy v roce 2005, RAM 4GB, rychlost dvou procesorů 3GHz na operačním systému Windows Server 2003 Standard<sup>28</sup>, databáze MS SQL 2005 – Singel<sup>29</sup>.

Poměrně velká časová náročnost je zapříčiněna především několikanásobnou kontrolou dat, výpočtem dne v týdnu a vyhledáváním textu v položce 8 struktury logu „D“, kde skript zjišťuje název použitého modulu a obrazovky, z které je uživatelem činěn dotaz.

---

<sup>28</sup> URL: < <http://technet.microsoft.com/en-us/windowsserver/bb512919>> [cit. 2011-12-12]

<sup>29</sup> URL: < <http://technet.microsoft.com/en-us/sqlserver/bb671245>> [cit. 2011-12-12]

## 4 Praktická část

Praktická část je rozdělena na tři základní bloky: získání, výpočet a zpracování informací pro řídicí, statistické a diagnostické účely. V jednotlivých blocích budou prováděny selekce relevantních dat, jejich zpracování a tvorba návrhu na uplatnění v praxi.

### 4.1 Využití dat z LogSouboru k řídicím účelům

Pro řídicí pracovníky jsou důležité především informace o geografickém rozložení sil a prostředků, časové využití pracovní doby, zatíženost informačního systému, dodržování normativ provozu, priorita jednotlivých modulů a vytížení uživatelů.

#### 4.1.1 Porovnání počtu cizinců a zatížení CIS podle spádových oblastí

Rozložení sil a prostředků je jedna z nejrelevantnějších informací pro řídicí pracovníky s celorepublikovou působností cizinecké policie. V prvním výpočtu porovnáme procentuální zastoupení cizinců žijících dlouhodobě na území České republiky v jednotlivých spádových oblastech a využívání modulů CIS pracovníky jednotlivých spádových oblastí.

##### *Výběr dat*

Geografické rozložení sil a prostředků lze ze zpracovaných dat zjistit podle IP adresy (Internet Protocol version 4, popsáno RFC 791)<sup>30</sup>, která je uchována v položce cLogIpD tabulky tLog v databázi „simi“. Samotný dotaz vybírá z tabulky jednotlivé IP adresy počítačů a u každé sčítá její výsky, a podle rozsahu IP adres přičítá dotazy jednotlivým spádovým oblastem.

---

<sup>30</sup> URL: <<http://tools.ietf.org/html/rfc791>> [cit. 2011-12-12]



### Dotaz SQL do tabulky tLog

```
use simi
SELECT cLogIpD, COUNT(*) AS CELKEM,
      SUM(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)<120 then 1
end, 0)) 'Praha',
      SUM(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>127 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<141 then 1 end, 0)) 'StC',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>143 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<152 then 1 end, 0)) 'JC',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>159 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<172 then 1 end, 0)) 'ZC',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>175 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<189 then 1 end, 0)) 'SC',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>191 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<205 then 1 end, 0)) 'VC',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>207 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<222 then 1 end, 0)) 'JM',
      SUM (ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>223 AND
CAST(SUBSTRING (cLogIpD,4,3) AS real)<245 then 1 end, 0)) 'SM'
from tLog
group by cLogIpD
order by cLogIpD
```

Výsledkem tohoto dotazu je tabulka, která u každé IP adresy disponuje celkovým součtem dotazů za rok 2009 a 2010 s označením příslušnosti ke spádové oblasti. Prostým součtem jednotlivých sloupců s počtem dotazů ve spádových oblastech dostaneme první část tabulky č. 5, ve které jsou v druhém sloupci uvedeny absolutní hodnoty počtu dotazů v jednotlivých spádových oblastech v celé České republice za roky 2009 a 2010. Vzhledem k tomu, že výsledná tabulka z SQL dotazu obsahuje IP adresy všech přístupujících PC a její velikost je 5500 řádků, není v diplomové práci uváděna. Uvedeny jsou jen konečné výpočty podle spádových oblastí.

Tabulka 5 - porovnání spádových oblastí dle cizinců a zatíženosti CIS

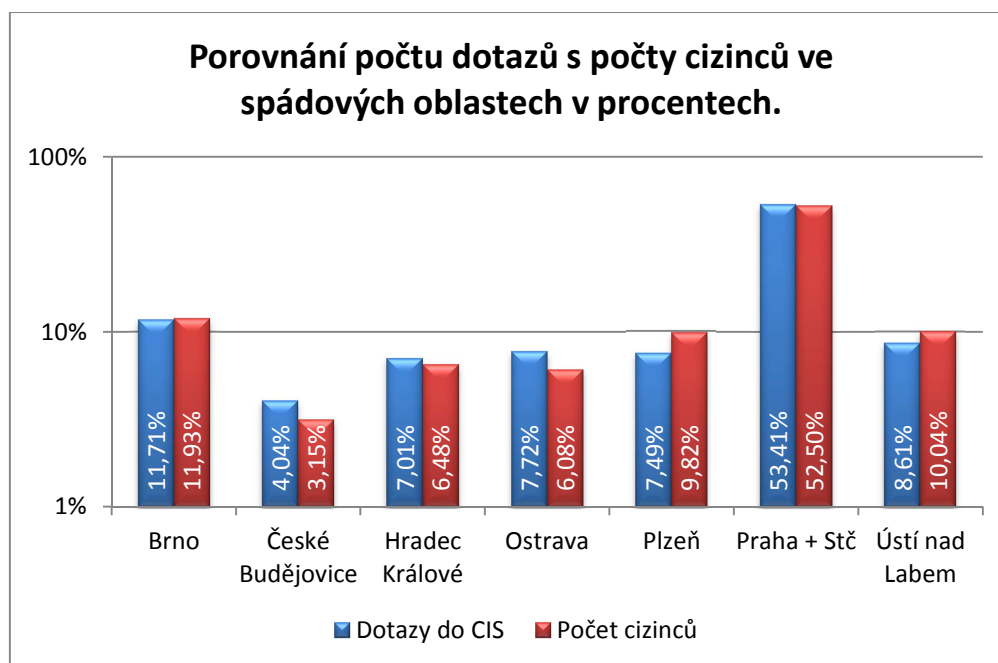
Spádová oblast	Dotazy do CIS		Počet cizinců(09+10)	
Brno	6190707	11,71%	58179	11,93%
České Budějovice	2136468	4,04%	15353	3,15%
Hradec Králové	3703085	7,01%	31606	6,48%
Ostrava	4077310	7,72%	29621	6,08%
Plzeň	3960159	7,49%	47863	9,82%
Praha + Stč	28227922	53,41%	255927	52,50%
Ústí nad Labem	4551864	8,61%	48934	10,04%
Celkem	52847515	100,00%	487483	100,00%

## Vnější data

Počty dotazů jsou porovnávány s počty cizinců na území spádových oblastí. Údaje o těchto počtech nejsou nikterak zadokumentovány v LogSouboru CIS. Z tohoto důvodu byly data převzaty z internetových stránek Ředitelství služby cizinecké policie, odboru analýzy rizik.<sup>31</sup>

## Použité metody a výpočty

Použity byly základní statistické metody relativní četnosti, které po vynásobení stem, udávají procentuální zastoupení v dotazech do CIS a počtu cizinců na území. Při výběru dat z databáze byl použit základní dotazovací příkaz SELECT. Data jsou porovnána v sloupcovém grafu (obrázek č. 6) s logaritmickým měřítkem. Logaritmické měřítko je vybráno z důvodu velké nesouměrnosti hodnoty Praha+Stč, v níž hodnoty přesahují 50%, ostatní údaje se pohybují řádově pod 10%.



Obrázek 6 – Graf porovnání počtu dotazů s počty cizinců ve spádových oblastech v procentech

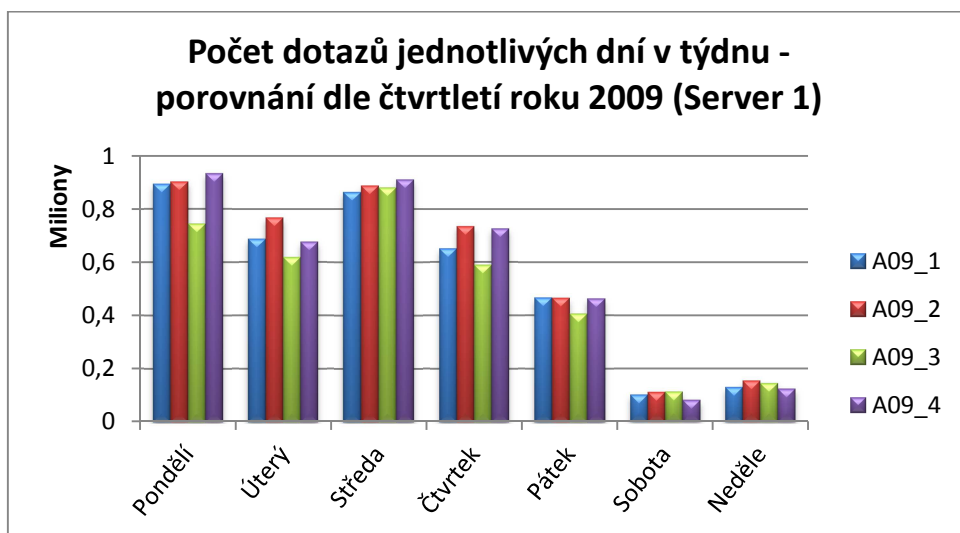
<sup>31</sup> URL: <<http://www.policie.cz/clanek/statistiky-885151.aspx>> [cit. 2011-12-12]

### *Forma uplatnění v praxi*

Na základě porovnání dvou údajů řídicí pracovník na celorepublikové úrovni získá přehled o vynaloženém úsilí pracovníků jednotlivých krajů k odbavení poměrného počtu cizinců. Pracoviště s procentuální odchylkou poměrné části odbavených cizinců k poměrné části provedených dotazů do informačního systému buď nadbytečně provádí některé úkony, nebo některé povinné úkony nevykonávají. Tento jev může být zapříčiněn i objektivní překážkou v činnosti útvarů. Tento druh překážky lze dalším rozbořem činnosti inkriminovaných útvarů odhalit a zaměřit se na její odstranění.

Výpočet tohoto porovnání byl proveden s daty za rok 2009 a 2010. Pro řídicí činnost se předpokládá měsíční nebo čtvrtletní výstup. V případě upřesnění rozsahu IP adres lze tento výstup uplatňovat i na konkrétní útvary, a tím dát k dispozici vedoucím nižších organizačních článků k dohledu nad svými podřízenými. Provedeným výpočtem v autentických datech cizinecké policie bylo zjištěno, že největší rozdíl je ve spádové oblasti Plzeň a činí přes dvě procenta. Naopak nejvíce vyrovnaný poměr je u spádových oblastí Praha+Stč a Brno.

Vypovídající hodnota porovnání procentuálního zastoupení dotazů a počtu odbavených cizinců žijících na teritoriu spádové oblasti může být zkreslena náhlým úbytkem nebo příbytkem cizinců v jednotlivých spádových oblastech.



Obrázek 7 - Graf porovnání počtu dotazů jednotlivých dní v týdnu

## 4.1.2 Hodinové zatížení serverů dotazy v roce 2009 a 2010

Mezi nejrelevantnější využití dat z bezpečnostního LogSouboru CIS je hodinové zatížení serverů. Hlavním principem tohoto výběru dat jsou hodinové četnosti dotazů na jednotlivé servery po dobu dvou let. Vzhledem k tomu, že výsledná tabulka disponuje údaji o každé hodině po dobu dvou let, lze výsledné sestavy zaměřit na kterýkoliv časový úsek v uvedené době. Konkrétní využití bude předvedeno v následujících odstavcích a kapitolách, jelikož hodnoty z výsledné tabulky budou využity i k analytickým a diagnostickým výpočtům.

### *Výběr dat*

Výběr dat je založen na krátkém dotazu příkazu SELECT, který počítá výskyt dotazu na serveru A a serveru B a seskupuje funkcí GROUP BY podle jednotlivých dní a hodin. Použitý dotaz dále sčítá pomocí SUM(ISNULL... . Počet dotazů na oba aplikační servery, jak bude vidět ve vygenerovaných tabulkách.

### *Dotaz SQL do tabulky tLog*

```
use simi
SELECT den=RTRIM(CONVERT(CHAR,dLogDateTimeD,102))+ ' '+CONVERT(
CHAR,DATEPART(hh, dLogDateTimeD)) ,
COUNT(*) AS Celkem,
SUM(ISNULL(case when bLogServer = 0 then 1 end, 0)) 'A',
SUM(ISNULL(case when bLogServer = 1 then 1 end, 0)) 'B'
from tLog

group by RTRIM(CONVERT(CHAR,dLogDateTimeD,102))+ ' '+CONVERT(
CHAR,DATEPART(hh, dLogDateTimeD))

order by den
```

Výsledkem dotazu je tabulka s počty dotazů v jednotlivých hodinách. Počty jsou dále řazeny podle serverů (značeny jako Server1 a Server2), na kterých je dotaz realizován. Vzhledem k rozsáhlosti tabulky je uvedena tabulka částečná s hodnotami prvních hodin v lednu roku 2009 a posledních hodin v prosinci roku 2010.

### Zjištění výskytu chyby

Podle známých údajů by měla mít tabulka celkem 17 520 řádků. Výpočet řádků je odvozen od počtu dní v roce násobeným dvěma a dále vynásobeným počtem hodin v jednom dni. Výpočet tedy vypadá takto:  $(365*2)*24=17\ 520$

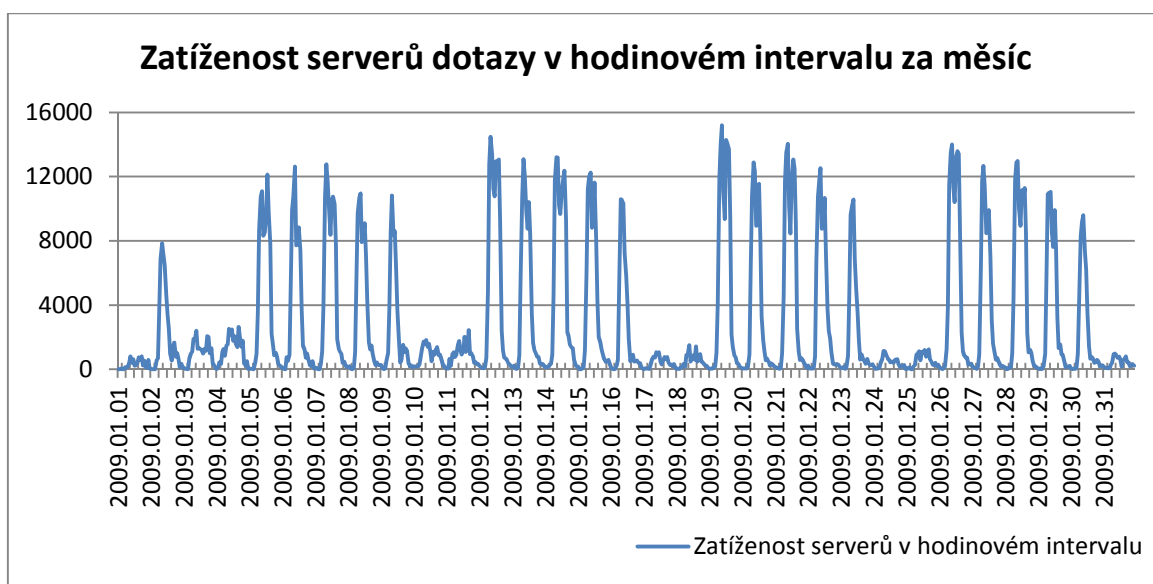
Tabulka 6 - hodinové četnosti dotazu na servery

Datum hodina	Celkem	Server 1	Server2	Pomocné data	
2009.01.01 0	0	0	0	2009.01.01	0
2009.01.01 1	9	0	9	2009.01.01	1
2009.01.01 2	10	0	10	2009.01.01	2
2009.01.01 3	52	0	52	2009.01.01	3
2009.01.01 4	4	0	4	2009.01.01	4
-----	-----	-----	-----	-----	--
2010.12.31 13	916	377	539	2010.12.31	13
2010.12.31 14	1503	693	810	2010.12.31	14
2010.12.31 15	1391	481	910	2010.12.31	15
2010.12.31 16	950	344	606	2010.12.31	16
2010.12.31 17	545	39	506	2010.12.31	17
2010.12.31 18	393	126	267	2010.12.31	18
2010.12.31 19	48	22	26	2010.12.31	19
2010.12.31 20	14	14	0	2010.12.31	20
2010.12.31 21	19	9	10	2010.12.31	21
2010.12.31 22	9	5	4	2010.12.31	22
2010.12.31 23	228	152	76	2010.12.31	23

Po výběru dat z databáze vznikla tabulka, která měla jen 17 452 řádků. Dalším šetřením v databázi a v tabulkách bylo zjištěno, že se jedná o absenci hodin, ve kterých nebyla zaznamenána žádná četnost dotazu, tedy žádná aktivita uživatelů. Pro odhalení těchto nesrovnalostí byla vytvořena kontingenční tabulka, kde v řádcích byl zadán datum a do pole výpočtů také datum. Tím vznikla další tabulka jednotlivých dní s počtem výskytu. Den, ve kterém nebyl výskyt 24, byl narušen nulovými hodnotami a opraven. Oprava spočívala v tom, že scházející hodina byla doplněna a vyplněna hodnotami „0“.

### ***Použité metody a výpočty***

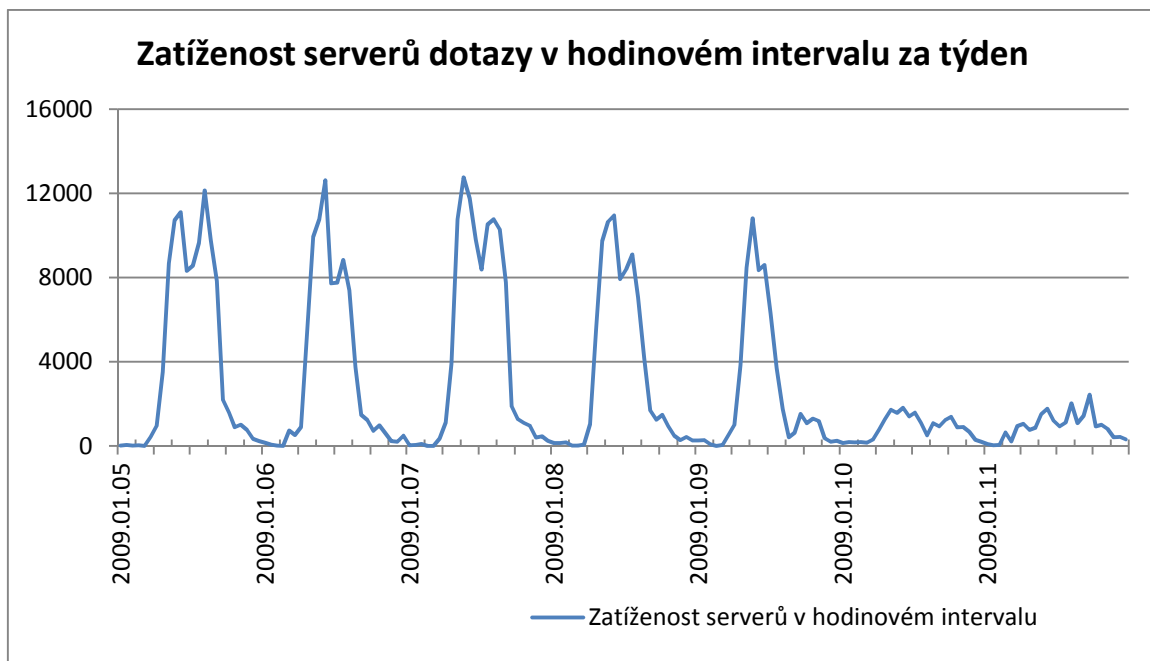
Použity byly základní statistické metody absolutní četnosti. Při výběru dat z databáze byl použit základní dotazovací příkaz SELECT s funkcí SUM, ISNULL, GROUPE BY, ORDER BY. K porovnání dat bylo využito spojnicového grafu (obrázek 8), který nevhodněji vystihuje časovou postoupnost zobrazených hodnot.



Obrázek 8 – Graf porovnání hodinového intervalu za měsíc

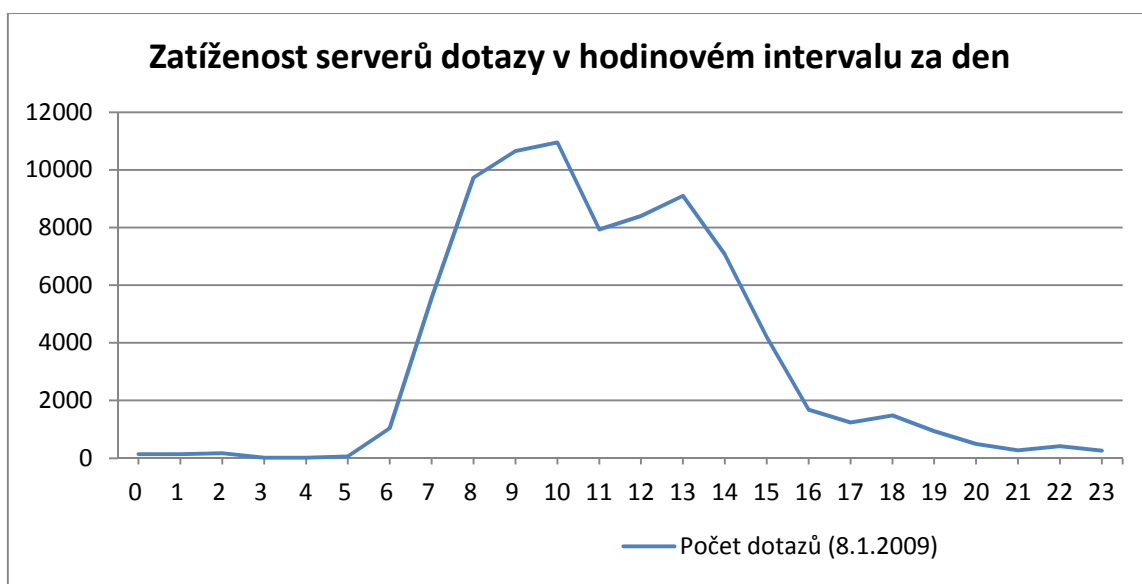
### ***Forma uplatnění v praxi***

První graf uvedený na obrázku číslo 8 vyhodnocuje počty dotazů za jednotlivé hodiny v měsíci leden 2009. Z tohoto pohledu lze zjistit pracovní nasazení v jednotlivých hodinách, dnech a celého měsíce v intervalu jedné hodiny.



Obrázek 9 – Graf porovnání hodinového intervalu dotazů za týden

Obdoba výstupu je zobrazena na obrázku číslo 9. Graf uvádí počtu dotazů jednotlivých hodin v týdnu. Při pohledu na tento graf jsou již patrné výkyvy v jednotlivých hodinách jednoho pracovního dne, kdy nasazení uživatelů není v předpokládaných osmi hodinách pracovní doby.



Obrázek 10 - Graf porovnání hodinového intervalu dotazů za den

Při využití dat s podrobností hodinového zatížení serverů v jednom dni (obrázek číslo 10) je již zcela patrné nasazení sil a prostředků, kdy po polední přestávce pracovní vytížení pravidelně klesá a již se nikdy nevrátí na hodnoty nejvyšší zatíženosti kolem deváté a desáté hodiny. Vypovídající hodnota porovnání jednotlivých hodin v měsíci, v týdnu a v jednom dni je značně přesná a její ovlivnitelnost objektivními překážkami je velice nepravděpodobná.

### 4.1.3 Porovnání plnění pokynu podle oblastí

Základním řídicím nástrojem je kontrola plnění pokynů podřízenými složkami. V CIS je vyžadován, až na výjimky, takzvaný důvod lustrace. Mezi důvody lustrace patří například kontrola osoby, žádost cizince o pobyt a podobně. Při transformaci z textové části LogSouboru do podoby relační databáze byl vypočítáván parametr „dotaz“, který v sobě uchovává text důvodu dotazu do CIS. Tento parametr byl zpracován do položky relační databáze pod název bLogDuvod jako binární data 0 nebo 1. Jednička značí zadaný dotaz do příslušné položky a nula značí dotaz bez udání důvodu. Následující výběr dat s grafickým výstupem porovná plnění tohoto pokynu v procentech podle jednotlivých oblastí.

#### *Výběr dat*

Geografické rozložení dotazů lze ze zpracovaných dat zjistit podle IP adresy (Internet Protokol version 4, popsáno RFC 791)<sup>32</sup>, která je uchována v položce cLogIpD tabulky tLog v databázi „simi“. Samotný dotaz vybírá z tabulky jednotlivé IP adresy počítačů a u každé sčítá její výskyt. Podle rozsahu IP adres přičítá dotazy jednotlivým spádovým oblastem a člení podle položky bLogDotaz na výskyty binárního znaku „0“ a „1“.

---

<sup>32</sup> URL: < <http://tools.ietf.org/html/rfc791> > [cit. 2011-12-12]



### *Doraz SQL do tabulky tLog*

```
use simi
select bLogDuvod, COUNT(*) AS CELKEM,

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)<120 then 1
end, 0)) 'Praha',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>127 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<141 then 1 end, 0)) 'StC',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>143 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<152 then 1 end, 0)) 'JC',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>159 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<172 then 1 end, 0)) 'ZC',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>175 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<189 then 1 end, 0)) 'SC',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>191 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<205 then 1 end, 0)) 'VC',

sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>207 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<222 then 1 end, 0)) 'JM',

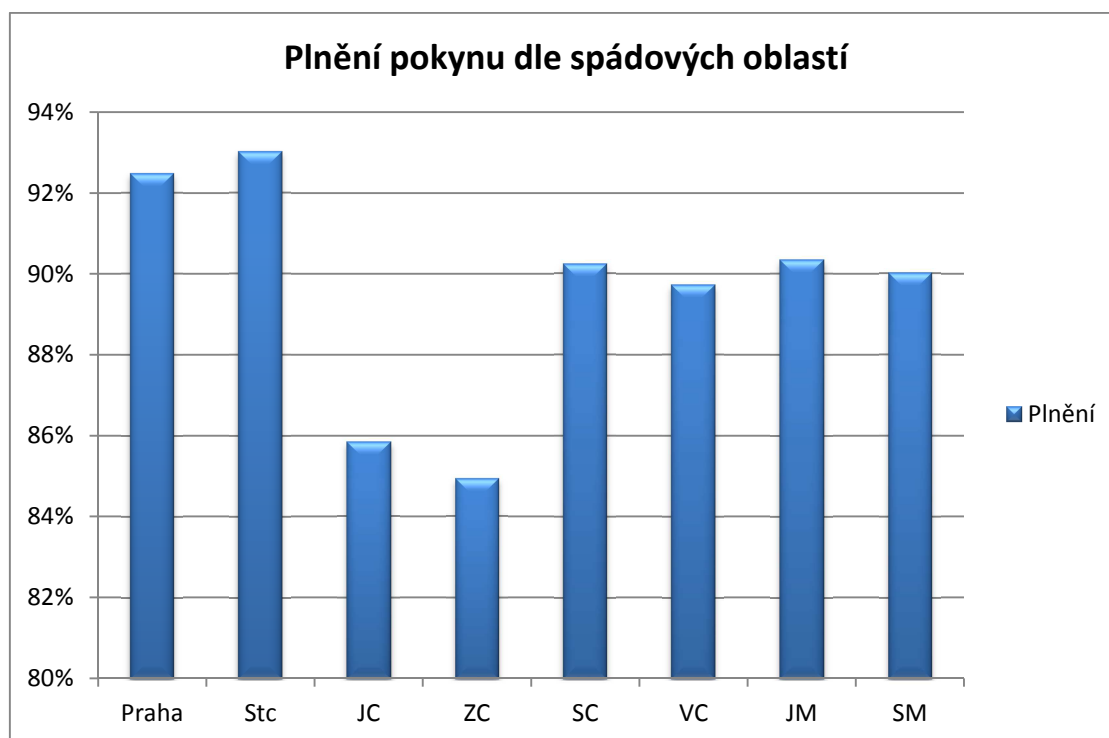
sum(ISNULL(case when CAST(SUBSTRING(cLogIpD,4,3) AS real)>223 AND
CAST(SUBSTRING(cLogIpD,4,3) AS real)<245 then 1 end, 0)) 'SM'
from tLog
group by bLogDuvod
```

Provedeným dotazem vznikne tabulka číslo 7, ve kterém jsou v prvním sloupci uvedeny jednotlivé kraje, v druhém sloupci absolutní četnost výskytu plnění jednotlivých krajů, v dalším sloupci absolutní četnost dotazů bez uvedení důvodu a v posledním sloupci relativní četnost v procentech.

**Tabulka 7- porovnání plnění pokynu zadávání důvodu dotazu dle krajů**

Kraj	Plnění pokynu	Neplnění pokynu	Plnění v %
<b>Praha</b>	22 684 456	1 705 886	92,48%
<b>Středočeský</b>	3 757 977	262 267	93,02%
<b>Jihočeský</b>	1 713 382	242 694	85,84%
<b>Západočeský</b>	3 425 258	516 118	84,93%
<b>Severočeský</b>	4 133 057	403 139	90,25%
<b>Východočeský</b>	3 347 224	343 690	89,73%
<b>Jihomoravský</b>	5 610 665	541 061	90,36%
<b>Severomoravský</b>	3 702 418	369 352	90,02%
<b>Celkem</b>	48 374 437	4 384 207	90,94%

Pro přehlednost jsou údaje k plnění pokynu graficky znázorněny v sloupcovém grafu (obrázek č. 11) s lineární stupnicí od 80% do 95%.



Obrázek 11 - Graf porovnání plnění pokynu dle spádových oblastí

### ***Forma uplatnění v praxi***

Informace v praxi především uplatní při své práci řídicí a kontrolní orgány pro další zjišťování objektivního nebo subjektivního důvodu neplnění pokynu nadřízených orgánů.

## 4.2 Využití dat z LogSouboru k statistickým účelům

Tato část diplomové práce je zaměřena na základní statistické údaje, které data bezpečnostního LogSouboru obsahují. Především se jedná o informace četnosti dotazů za využití jednotlivých modulů, obrazovek. Četnosti dotazů z jednotlivých oblastí, počtu standardních dotazů, počtu odpovědí, největší hodinové zatíženosti a podobně.

Informace z těchto výstupů využijí především pracovníci IT podpory PČR, kteří plánují odstávky systému, detekují nestandardní chování systému a vedoucí pracovníci základních útvarů k zjištění nestandardního nasazení sil a prostředků v konkrétním období. Jedná se především o konkrétní prudké růsty aktivity uživatelů, a nebo naopak jejich nečinnost.

Výběr dat a výpočty jednotlivých relevantních informací se omezí v zájmu rozsahu diplomové práce jen na tři ukázky s komentářem.

### 4.2.1 Základní statistické údaje z CIS

Mezi základní statistické údaje jsou zařazeny absolutní počty dotazů, odpovědí, jejich krajní meze, jakými jsou například MAX absolutní četnost dotazů za hodinu, MIN absolutní četnost za hodinu a podobně.

#### *Výběr dat*

K statistickým výpočtům bude převážně použit výběr z databáze v podobě tabulky č. 6 z podkapitoly 4.1.2. V rámci přehlednosti, omezení počtu tabulek, omezení počtu grafů a dodržení rozsahu této práce budou některé prvky a výpočty zveřejněny v rámci přílohy této práce.

**MAX**imální absolutní četnost dotazů za hodinu byla vypočtena u prvních deseti nejzatíženějších hodin v roce 2009 a 2010. Výsledek je v tabulce číslo 8, v níž autor uvádí postupné pořadí, název dne v týdnu, hodinu, datum, počet dotazů a průměrný počet dotazů za vteřinu. Uvedená tabulka vypovídá především o tom, že v největší zatíženosti serverů je zpracováváno od 6 do 7 dotazů za vteřinu. Mezi nejnáročnější dny vůči zatíženosti serverů

k počtu dotazů patří pondělí v době od 10 do 14 hodin. Posléze následuje středa v 10 hodin.

**Tabulka 8 - deset nejzatíženějších hodin dotazy obou serverů v roce 2009 a 2010**

Poř.	Den v týdnu	Hodina	Datum	Počet dotazů	Průměr na vteřinu
1	pondělí	10	9.11.2009	24141	6,7
2	pondělí	13	9.11.2009	21484	6,0
3	středa	10	11.11.2009	21357	5,9
4	pondělí	14	9.11.2009	21315	5,9
5	pondělí	10	30.11.2009	21310	5,9
6	pondělí	9	30.11.2009	20868	5,8
7	pondělí	13	2.11.2009	20822	5,8
8	úterý	10	3.11.2009	20419	5,7
9	středa	10	4.11.2009	20367	5,7
10	pondělí	14	2.11.2009	20016	5,6

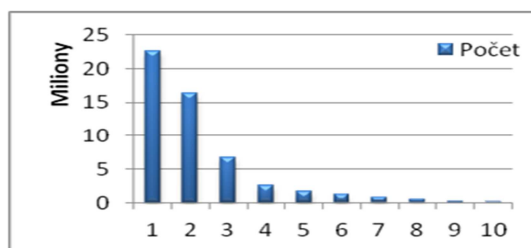
**MIN**imální absolutní četnost dotazů na jednu hodinu nelze tímto způsobem vypočítat, jelikož v době dvou let je celkem 68 hodin, ve kterých dotaz do Cizineckého informačního systému nebyl v hodinovém intervalu ani jeden. Z tohoto důvodu bylo přistoupeno k výpočtu četností podle jednotlivých dní v týdnu a jednotlivých hodin v konkrétním dni, kdy není učiněn ani jeden dotaz, jak je uvedeno v obou částech tabulky číslo 9. Údaje první části tabulky vypovídají o tom, že nejméně dotazů, respektive žádné dotazy, nejsou v neděli a ve čtvrtek mezi 3. a 4. hodinou ranní, celkem v 37 případech. Pokud se jedná jen o konkrétní hodiny v jednom dni, tak žádné dotazy nejsou celkem v 50 případech mezi 3. a 4. hodinou (pravá část tabulky číslo 9).

**Tabulka 9 - nečinnosti na aplikačních serverech v průběhu roku 2009 a 2010**

Dny v týdnu	Výskyt
Pondělí	9
Úterý	4
Středa	6
Čtvrtek	17
Pátek	7
Sobota	5
Neděle	20

Hodiny:	Výskyt
0	2
1	3
2	9
3	25
4	25
5	2
6	2

K ohodnocení souboru dat celkového zatížení serverů dotazy za dva roky v hodinovém intervalu bylo použito metody **KVANTIL**<sup>33</sup> s rozdělením na deset částí, to je **DECIL**. K samotnému výpočtu byl použit tabulkový editor MS Excel a jeho funkce PERCENTIL()<sup>34</sup> s hodnotami od 0,1 do 1 v intervalu 0,1, čímž bylo docíleno hodnot DECIL, které jsou uvedeny v posledním sloupci tabulky číslo 10. Pro lepší přehled zastoupení jednotlivých hodnot hodinového počtu dotazů v CIS byl proveden ještě výpočet průměru dotazů za jednu kvantilu, procentuální zastoupení (relativní četnost\*100) a kumulativní procentuální zastoupení.



Obrázek 12 – Graf pomocný

Uvedený přehled informuje o skutečnosti, že 91% dotazů do informačního systému CIS je uskutečněno v jedné čtvrtině celého dne, což je dobře vidět na pomocném grafu obrázku číslo 12, ve kterém jsou uvedeny absolutní hodnoty počtu dotazů v jednotlivých decilech. Ve zbylých třech čtvrtinách je uskutečněno pouze 19% dotazů.

Tabulka 10 - rozdělení základního souboru hodinového počtu dotazů

Pořadí	Počet	průměr	Průměr za min.	Zast. % pom.	Zast. %	DECIL
1	22 641 700	12923	215,38	42,92%	42,92%	24 141
2	16 438 509	9382,2	156,37	31,16%	74,07%	11 049
3	6 748 578	3851,7	64,20	12,79%	86,87%	7 131
4	2 537 073	1448	24,13	4,81%	91,67%	1 897
5	1 712 660	977,49	16,29	3,25%	94,92%	1 157
6	1 213 183	692,42	11,54	2,30%	97,22%	823
7	799 802	456,48	7,61	1,52%	98,74%	567
8	446 977	255,11	4,25	0,85%	99,58%	351
9	177 798	101,48	1,69	0,34%	99,92%	163
10	42 353	24,173	0,40	0,08%	100,00%	53

<sup>33</sup> SVATOŠOVÁ, Libuše – KÁBA, Bohumil.: Statistické metody I. Česká zemědělská univerzita v Praze. 2009. ISBN 978-80-213-1688-0.

<sup>34</sup> URL: < <http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx> > [cit. 2011-12-12]

Tato skutečnost je patrná i z prováděných výstupů v ostatních měřeních, která byla vypočítávána v kapitole 4.1.

### ***Forma uplatnění v praxi***

Provedený výstup představuje informace o době maximálního zatížení aplikačních serverů. Upřesňuje dny a hodiny, které jsou kritické na největší zatížení. V tomto časovém období je nejméně vhodné provádět komplikované dotazy do CIS, plánovat výpadek za účelem údržby a podobně. Z druhé strany, výstup nachází časové období v pracovním týdnu, kdy je aktivita uživatelů minimální až nulová. V tomto časovém období lze hledat optimální předpoklady pro náročné dotazy do systému, sehrávání dat, plánování údržby a odstávky systému.

Porovnáním základního souboru bylo zjištěno, že největší počet dotazů se uskutečňuje u dvou decilech dne, což lze převést na čtyři až pět hodin denně. Po tuto dobu je CIS zatížen nadměrně. „Špička“ dotazů na aplikační servery se zpracovává v jedné desetině dne, to je dvě a půl hodiny denně. Samotné největší zatížení lze redukovat rozptěním úkonů v CIS na celou osmihodinovou pracovní dobu.

## **4.2.2 Počty pohybu uživatelů v programovém prostředí CIS**

Mezi další informace, které lze získat z databáze „simi“ a její tabulky tLog, jsou počty využití jednotlivých modulů Cizineckého informačního systému, jejich obrazovek a druhů dotazu.

### ***Výběr dat***

K výběru dat z databázové tabulky tLog provede jednoduchý dotaz za využití příkazu SELECT a funkcí LEFT() a COUNT(), který vytvoří tabulku s názvem modulu, jeho četností dotazu za rok 2009 a 2010. Řídící položkou je cLogMod, v němž je uchován plný název modulu. Následně bude vypočtena relativní četnost každého modulu zvlášť.

### *Dotaz SQL do tabulky tLog*

```
use simi
select LEFT(cLogMod,3) AS 'MODUL', COUNT(*) AS pocet from tLog
group by LEFT(cLogMod,3)
order by LEFT(cLogMod,3)
```

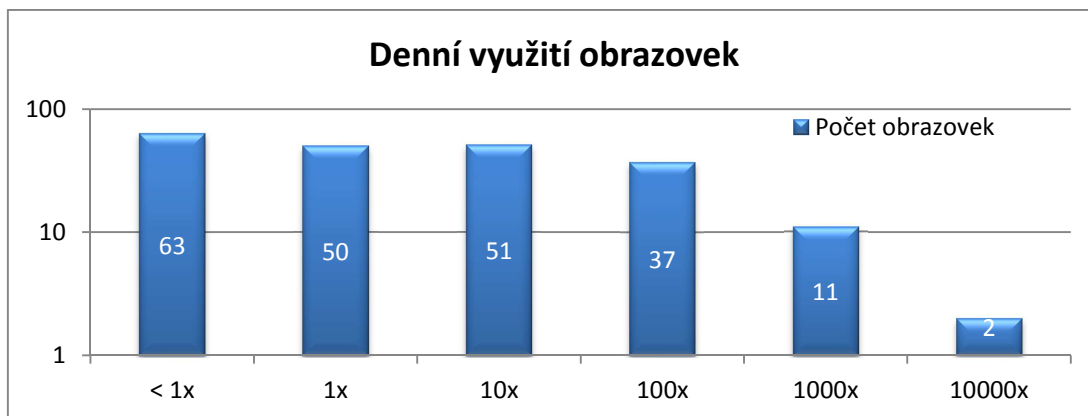
Modulů v CIS je celkem 28. Více jak dvě třetiny jsou v relativní četnosti výskytu za dva roky zanedbatelné. Tabulka proto obsahuje jen prvních deset nejvíce užívaných modulů.

**Tabulka 11 - počet dotazů do jednotlivých modulů CIS**

č	Modul	AbsolutČetnost	RelativČetnost
1	TDU	32874962	62,15%
2	PLU	6597422	12,47%
3	LUS	4231142	8,00%
4	EVIC2	3532049	6,68%
5	UBY	2043350	3,86%
6	VIZ	903471	1,71%
7	POZ	901814	1,70%
8	PRE	342591	0,65%
9	AZYL	336351	0,64%
10	SPS	323618	0,61%

I bez grafického znázornění lze z tabulky vyčíst absolutní převahu využití modulu „Trvale a dlouhodobě ubytovaní cizinci“ (TDU), kde relativní četnost ukazuje 62,15% z celkového počtu dotazů za dva roky na oba aplikační servery. Dále následují moduly „Podrobná lustrace „ (PLU) s 12,47% a „Rychlá lustrace“ (LUS) s 8,00%.

**Výběr dat** k položce cLogObr a cLogDruh a jejich dotaz SQL je uveden v příloze číslo 1. V této části práce jsou zobrazeny jen výsledné tabulky grafy s jejich interpretací.



Obrázek 13 - Graf porovnání denního využití obrazovek

V případě výstupu počtu dotazů z obrazovek (pojmenovaného grafického prostředí CIS) byl proveden výpočet četnosti dotazů na jeden typ obrazovky. Více jak 10 000 x za den jsou využity dvě obrazovky, a to TDU-Detail osoby a TDU-Vyhledávání osoby, jak je vidět na grafu obrázku číslo 13. Více jak 1000x za den je použito dalších 11 obrazovek. Třetí v pořadí je obrazovka PLU-Podrobná lustrace, která je průměrně využita 6.004 krát za den.

**Druh dotazu** je další informace o pohybu uživatele v systému. Druhem dotazu je myšleno působení uživatele na databázi, tudíž zápis do databáze, výběr dat nebo editace.

Druh dotazu byl zpracován obdobným způsobem jako četnost denního využití obrazovek. Výpočtem bylo zjištěno, že největší četnost dotazů je ve formě výběru dat (lustrace) s počtem 43 289 441 za období roku 2009 a roku 2010. Následuje zápis dat do databáze s počtem 5 987 435, změna dat v databázi s počtem 3 568 818 a výmaz dat s počtem 62 093.

Z uvedených informací je získán přehled o hlavní činnosti uživatelů, kteří především vyhledávají data v databázi k dalšímu zpracování. Tato činnost je dominantní, jelikož téměř desetinásobně převyšuje počet úkonů na druhém místě, což je zápis do databáze. Obecně lze označit, že k jednomu zápisu do databáze je potřeba deseti lustrací.

**Relevantnost výstupu** a interpretace je ovlivněna rozličnou náplní práce složek cizinecké policie, která k svoji práci využívá Cizinecký informační systém.



### 4.2.3 Porovnání časové náročnosti jednotlivých modulů

Zjištění časové náročnosti na zpracování dotazu jednotlivých modulů lze optimalizovat zadáním jednotlivých modulů, využití restrikce pro uživatele u náročných operací, případně zadat autorovi sjednání nápravy. Tento výstup je více určen pro diagnostické účely, než pro statistické. V zájmu sledu provedených výstupů je ovšem zařazen do této kategorie.

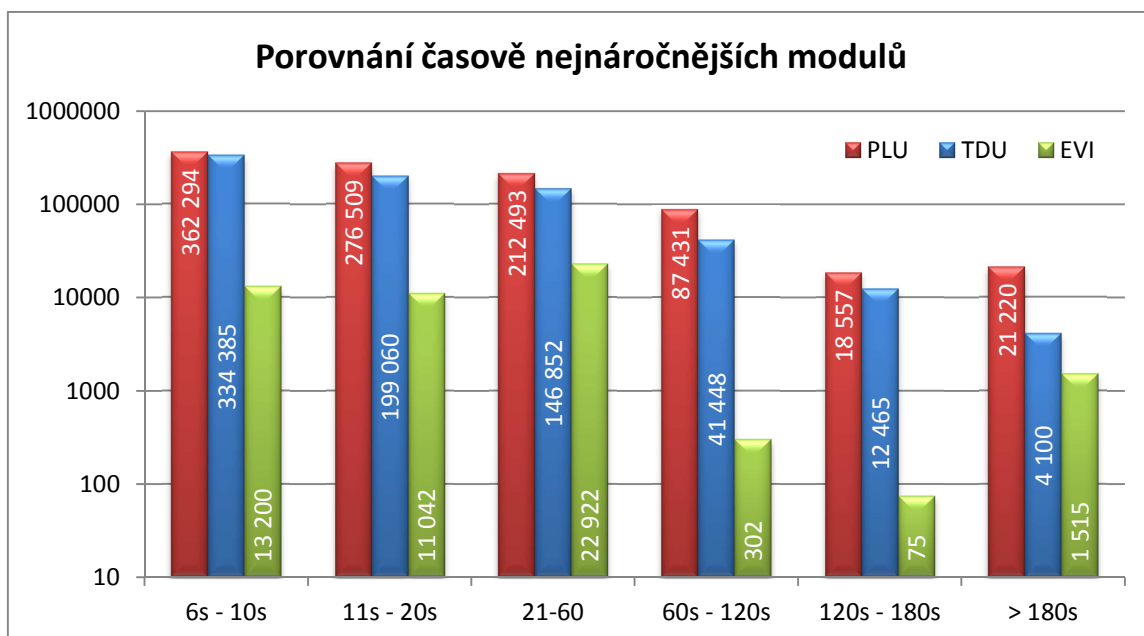
#### *Výběr dat*

K výběru dat z databázové tabulky tLog je proveden jednoduchý dotaz za využití příkazu SELECT a funkcí LEFT() a COUNT(), který vytvoří tabulku s názvem modulu a četností jeho dotazu za rok 2009 a 2010. Řídící položkou je nLogDelkaZprac, ve které je uchován plný název modulu. Následně bude vypočtena relativní četnost každého modulu zvlášť s délkami zpracování dotazu od vteřin do 3 minut.

#### *Dotaz SQL do tabulky tLog*

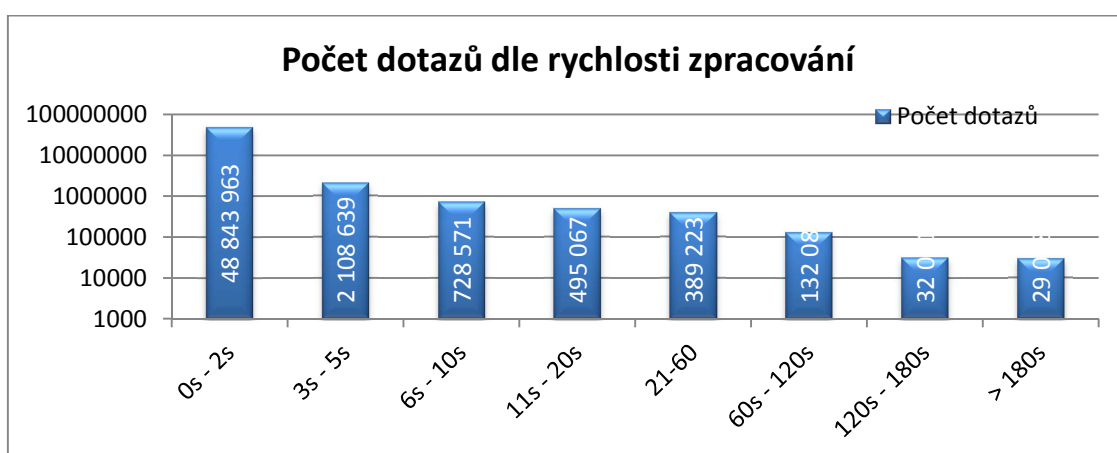
Pro svoji rozsáhlost je dotaz SQL do tabulky tLog uveřejněn v příloze číslo 2.

Výsledkem dotazu je tabulka s výpisem jednotlivých modulů, jejich celkovou četností a četností podle času s intervalem od jednotek vteřin do třech minut. Tabulka je součástí přílohy číslo 2. Z tabulky byl vytvořen sloupcový graf s logaritmickou stupnicí, v níž jsou patrné rozdíly četnosti dotazů z jednotlivých modulů v návaznosti na dobu trvání dotazu. Do doby trvání byly vybrány četnosti, které nejsou v normě Cizineckého informačního systému. Jedná se o čas odezvy delší jak 5 vteřin.



Obrázek 14 - Graf porovnání časově nejnáročnějších modulů

Výstupem je sloupcový graf s logaritmickou stupnicí, který uvádí celkové počty dotazů v letech 2009-2010 s rozdělením doby provedení dotazu podle stupnice znázorněné pod horizontální osou. Logaritmická stupnice vertikální osy je vybrána z důvodu velkého číselného rozsahu jednotlivých údajů. Porovnání tří nejpoužívanějších modulů je v grafu na obrázku číslo 14.



Obrázek 15 – Graf porovnání počtu dotazů dle rychlosti zpracování.

Cizinecký informační systém garantuje odezvu dotazu do 5 vteřin. Celkovým počtům je věnován další výstup, který je znázorněn na grafu v obrázku číslo 15. První sloupec vyznačuje četnost dotazů provedených do 2 vteřin. Jejich absolutní četnost je 48 843 693, což je 92,55% z celkového počtu, jak je patrné z grafu obrázku číslo 15. Druhý sloupec vyznačuje četnost dotazů v toleranci od 3 do 5 vteřin. Jejich absolutní četnost je 3 566 724, což činí 4,01%. Hodnoty těchto čísel dokazují, že Cizinecký informační systém v 96,56% vyhovuje požadavkům zadavatele zakázky a v drtivé většině dotazů je rychlost do dvou vteřin. Zbýlých 3,44% dotazů si vyžaduje hlubší analýzu za účelem zjištění příčin nedodržení stanovené lhůty. Mezi těmito údaji jsou například časy při sehrávání dat, zálohování, analytické dotazy, nekorektní chod aplikačního serveru a podobně.

**Relevantnost výstupu** není při výběru dat a dalších výpočtech subjektivně ani objektivně ovlivněna. Samotná doba zpracování dotazů může být ovlivněna neznalostí uživatelů s prací modulu Podrobná lustrace, který umožňuje zadat dotaz, jenž není plně optimalizovaný.

#### **4.2.4 Základní statistické údaje o vygenerovaných odpovědích**

Výběr dat a propočty jsou do současné doby prováděny z dat četnosti dotazů. Každý dotaz zaslaný aplikačnímu serveru čeká na odpověď, kterou vygeneruje programové vybavení serveru. Další část statistických údajů se bude týkat vygenerovaných odpovědí na dotazy zaslané klientem.

##### ***Výběr dat***

Popis SQL dotazu je v zájmu úspory místa v hlavní části práce v příloze číslo 3.

**Celkový počet odpovědí** vygenerovaných oběma aplikačními servery v průběhu let 2009 a 2010 je **710 667 044**. V přepočtu průměru na jeden den se jedná o 40 563 odeslaných řádků odpovědí a něco přes 12 odpovědí za vteřinu. Samotná uvedená čísla nedisponují relevantní informací, pokud nebudou porovnávána se samotnými dotazy. K tomuto účelu byla zkonstruována tabulka číslo 12.

Tabulka 12 - porovnání optimalizovaných a neoptimalizovaných dotazů

Počet odpovědí	Počet výskytu	Za den	Zast. v %	Optimal. dotaz	Standard dotaz
0	7 483 542	10251	14,18%	94,27%	98,28%
1_10	42 253 156	57881	80,09%		
11_20	1 147 369	1572	2,17%	5,73%	1,72%
21_50	969 852	1329	1,84%		
51_100	509 760	698	0,97%		
101_500	339 653	465	0,64%		
500_1K	28 353	39	0,05%		
1K_10K	22 492	31	0,04%		
10K_100K	3 693	5	0,01%		
100K_1M	1 418	2	0,00%		
1M_10M	40	0	0,00%		
<b>Celkem:</b>	52 759 328	72 273	100%	100%	100%

V prvním sloupci jsou uvedeny intervaly počtu vygenerovaných odpovědí na jeden dotaz. Ve skutečnosti se jedná například o počet vrácených identit cizinců, které byly dotazem zpracovány a jejich kolekce vrácena zpět klientovi k dalšímu použití. Za optimalizovaný dotaz je považován ten, který vrátí odpovědi v rozmezí od 0 do 10 řádků v kolekci. Jako standardní je považován dotaz, který vrátí od 11 do 50 odpovědí. Ostatní dotazy, které vrátí více, než 50 odpovědí, jsou považovány za nestandardní.

Na základě výběru všech dotazů a jejich počtu odpovědí v průběhu let 2009-2010 bylo zjištěno, že z 94,27% byly zadány dotazy zcela optimalizované. 98,28% z celkového počtu dotazů se jedná o dotazy optimalizované nebo standardizované. Celkem v 1,72% případech dotazů se jedná o dotazy nestandardizované. Dalším rozбором nestandardizovaných dotazů by bylo možné dojít až k příčině důvodu, to ovšem není předmětem této práce.

**Relevantnost výstupu** není při výběru dat a dalších výpočtech subjektivně ani objektivně ovlivněna. Samotná nestandardnost dotazů může být ovlivněna neznalostí uživatelů s prací modulu Podrobná lustrace, který umožňuje zadat dotaz, jenž není plně optimalizovaný.

## 4.3 Využití dat z LogSouboru k diagnostickým účelům

Na první pohled je dosti nelogické, aby pracovníci Policie České republiky (PČR) prováděli diagnostiku informačního systému, který udržuje a odpovídá za chod autorská soukromá firma. Jsou to ale právě pracovníci IT podpory PČR, kteří následovně komunikují s technikou a autory díla CIS. Při každém návrhu z řad uživatelů na změnu programového provozního prostředí je potřeba fundovaně rozeznat potřeby funkčně relevantní a ekonomicky dostupné. Bez vlastních diagnostických nástrojů a vědomostí nelze komunikovat s firmou xy.cz na odborné úrovni v rámci finančně nákladové šetrnosti a udržitelnosti chodu systému v maximálním uživatelském komfortu.

### 4.3.1 Porovnání vytíženosti aplikačních serverů v hodinovém intervalu

Mezi nejrelevantnější využití dat z bezpečnostního LogSouboru CIS rozhodně patří hodinové zatížení serverů. Hlavním principem tohoto výběru dat jsou hodinové četnosti dotazů na jednotlivé servery po dobu dvou let. Vzhledem k tomu, že výsledná tabulka disponuje údaji o každé hodině dvouletého období, lze výsledné sestavy zaměřit na kterýkoliv časový úsek v uvedené době.

#### *Výběr dat*

Výběr dat je založen na krátkém dotazu příkazu SELECT, který počítá výskyt dotazu na serveru A a serveru B, kde seskupuje funkcí GROUP BY jednotlivé dny. Použitý dotaz dále sečte pomocí SUM(ISNULL... . počet dotazů pro oba aplikační servery, jak je zobrazeno ve vygenerovaných tabulkách, které autor uvádí v kapitole 4.2.2.

### *Dotaz SQL do tabulky tLog*

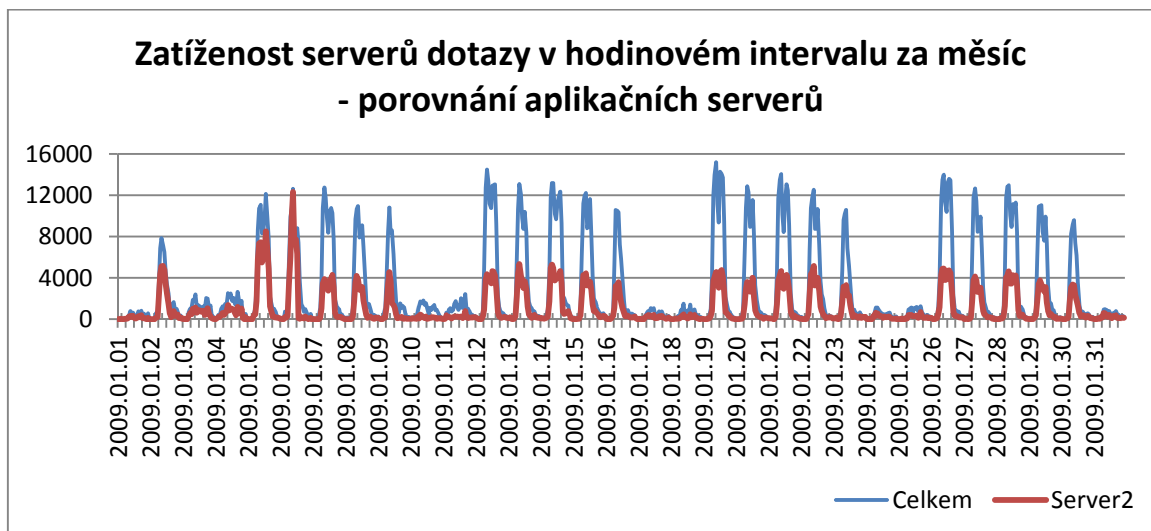
```
use simi
SELECT den=RTRIM(CONVERT(CHAR,dLogDateTimeD,102))+ ' '+CONVERT(
CHAR,DATEPART(hh, dLogDateTimeD)) , COUNT(*) AS Celkem,
SUM(ISNULL(case when bLogServer = 0 then 1 end, 0)) 'A',
SUM(ISNULL(case when bLogServer = 1 then 1 end, 0)) 'B'
from tLog
group by RTRIM(CONVERT(CHAR,dLogDateTimeD,102))+ ' '+CONVERT(
CHAR,DATEPART(hh, dLogDateTimeD))

order by den
```

Výsledkem dotazu je tabulka s počty dotazů v jednotlivých hodinách. Počty jsou dále řazeny podle serverů značeny jako Server1 a Server2 na kterých je dotaz realizován. Tabulka číslo 2 je uvedena v kapitole 4.1.2 na straně 39.

### *Použité metody a výpočty*

Použity byly základní statistické metody seskupování a absolutní četnosti. Při výběru dat z databáze byl použit základní dotazovací příkaz SELECT s funkcí SUM, ISNULL, GROUPE BY, ORDER BY. K porovnání dat je využito spojnicového grafu, který nevhodněji vystihuje časovou postoupnost zobrazených hodnot.

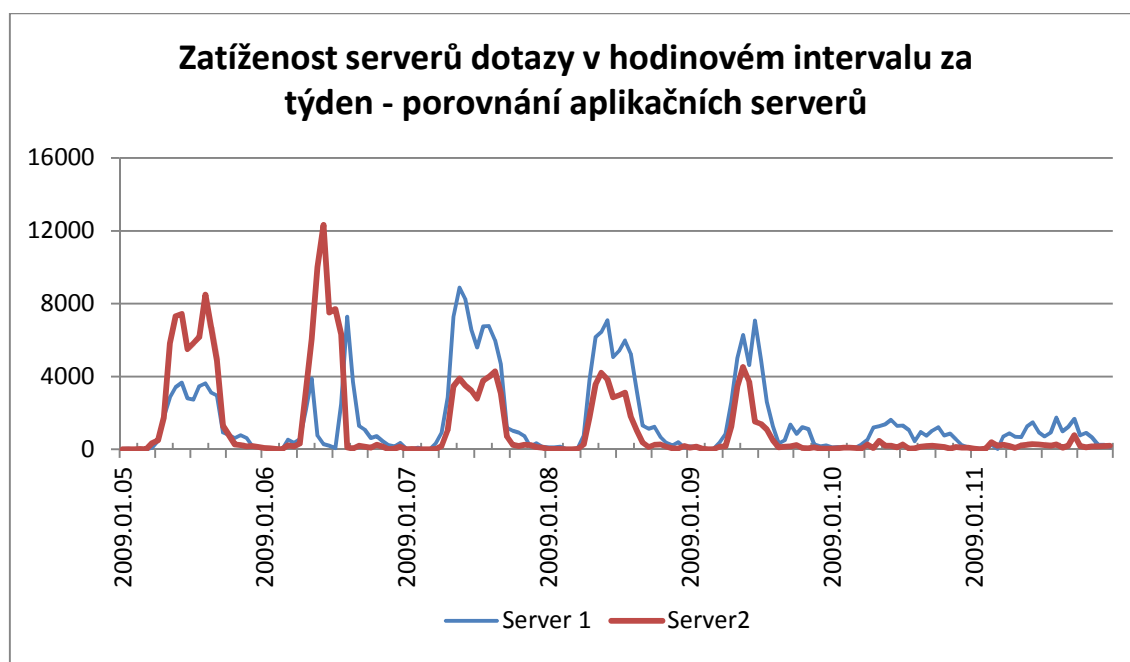


Obrázek 16 - Graf porovnání aplikačních serverů za měsíc

U grafu zobrazeném na obrázku č. 16 je využito rozdílu hodnot celkových a Serveru 2. Modrá část (hodnota Celkem) nad červenou částí Serveru 2 grafu zobrazuje zbylé hodnoty, tedy hodnoty Serveru 1. Tímto způsobem je dosaženo efektu porovnání obou serverů, jelikož server 1 je vyobrazen (modrá barva křivky) jako rozdíl celkového součtu a Serveru 2.

### *Vypovídající hodnoty*

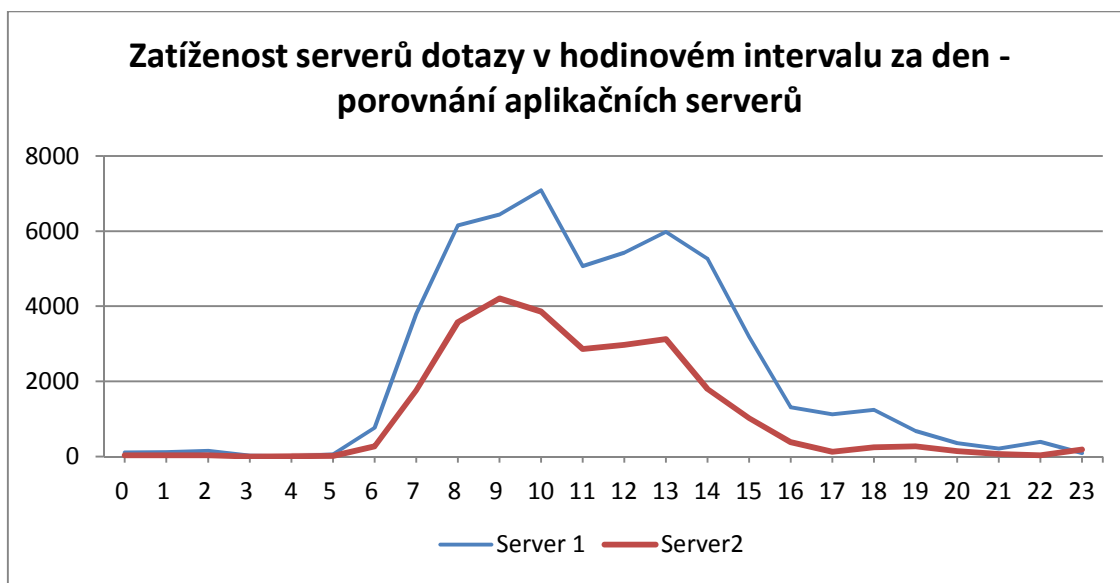
První graf uvedený na obrázku číslo 16 vyhodnocuje počty dotazů za jednotlivé hodiny v měsíci leden 2009 u Serveru1 s porovnáním počtu dotazů za uvedené období Serveru2. Z tohoto pohledu lze zjistit zatíženost obou aplikačních serverů v jednotlivých hodinách, dnech a celém měsíci v intervalu jedné hodiny.



**Obrázek 17 - Graf porovnání aplikačních serverů za týden**

Obdoba výstupu je zobrazena na obrázku číslo 17, kde je graf počtu dotazů jednotlivých hodin v týdnu. Při pohledu na tento graf jsou již patrné výkyvy v jednotlivých hodinách jednoho pracovního dne, kdy zatíženost aplikačních serverů není stejná. Uprostřed dne 6.1.2009 je **znatelná kolize** Serveru1, jehož počet dotazů klesl na 0, a počet dotazů

Serveru2 prudce stoupl přes 12 000 za hodinu. Při využití dat s podrobností hodinového zatížení serverů v jednom dni, jak zobrazuje graf na obrázku č. 18, je již zcela patrný rozdíl počtu dotazů na Serveru1 a Serveru2. Tento **jev nesouměrnosti počtu dotazů** na oba aplikační servery se vyskytuje v celém období.



Obrázek 18 - Graf porovnání aplikačních serverů za den

Dále bylo přistoupeno k výpočtu porovnání závislosti počtu dotazů na Serveru1 a Serveru2. K výpočtu autor použil korelační koeficient<sup>35</sup>, který byl vypočítán z údajů uvedené tabulky, v níž jsou porovnána jednotlivá hodinová data Serveru1 a Serveru2 za období dvou let. K výpočtu byl použit tabulkový editor MS Excel a jeho funkce CORREL()<sup>36</sup>. Provedený výpočet dosáhl hodnoty 0,48545492. Tato hodnota ukazuje přímou střední závislost počtu dotazů zpracovaných na Server1 a Server2. Podrobná analýza uvedeného jevu není předmětem této práce.

**Vypovídající hodnota** porovnání jednotlivých hodin v měsíci, v týdnu a v jednom dni je dostatečně přesná a její ovlivnitelnost objektivními překážkami je velice nepravděpodobná.

<sup>35</sup> SVATOŠOVÁ, Libuše – KÁBA, Bohumil.: Statistické metody I. Česká zemědělská univerzita v Praze. 2009. ISBN 978-80-213-1688-0.

<sup>36</sup> URL: < <http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx> > [cit. 2011-12-12]



### 4.3.2 Porovnání funkčnosti „BalancerServeru“

Mezi významné aktivní prvky Cizineckého informačního systému patří BalancerServer<sup>37</sup>, který řídí zátěž obou aplikačních serverů a v případě chybovosti odkloní zátěž na plně funkční stroj. V zcela ideálních podmínkách by měla být zátěž obou serverů stejná. Pomyslnou rovnováhu obou serverů zjistí výstup, který porovná počet zpracovaných dotazů za dva roky s měsíčním intervalem na obou aplikačních serverech.

#### *Výběr dat*

Výběr dat je založen na dotazu příkazu SELECT, který vybírá výskyt dotazu na serveru A a serveru B. Posléze seskupuje data funkcí GROUP BY podle měsíců a dní a sčítá pomocí SUM(ISNULL.... počet dotazů na oba aplikační servery za roky 2009 – 2010.

#### *Dotaz SQL do tabulky tLog*

```
use simi
select MONTH(dLogDateTime) AS mesic, COUNT(*) AS pocet ,
SUM(ISNULL(case when YEAR(dLogDateTime)=2009 AND bLogServer=0 then 1
end, 0)) '09_1',
SUM(ISNULL(case when YEAR(dLogDateTime)=2009 AND bLogServer=1 then 1
end, 0)) '09_2',
SUM(ISNULL(case when YEAR(dLogDateTime)=2009 then 1 end, 0)) '09',
SUM(ISNULL(case when YEAR(dLogDateTime)=2010 AND bLogServer=0 then 1
end, 0)) '10_1',
SUM(ISNULL(case when YEAR(dLogDateTime)=2010 AND bLogServer=1 then 1
end, 0)) '10_2',
SUM(ISNULL(case when YEAR(dLogDateTime)=2010 then 1 end, 0)) '10'
from tLog
group by MONTH(dLogDateTime)
order by MONTH(dLogDateTime)
```

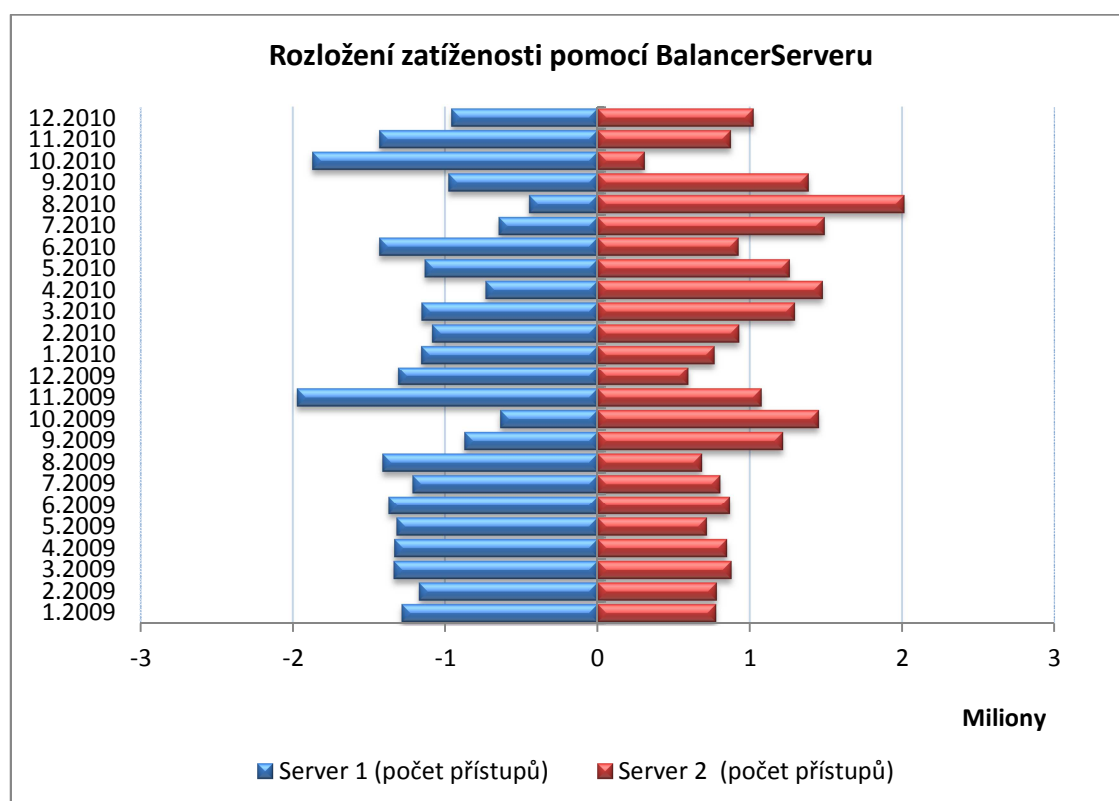
Výsledkem dotazu je tabulka s počty dotazů v jednotlivých měsících. Počty jsou dále děleny podle serverů (značeny jako Server1 a Server2) s rozdělením let 2009 a 2010. Tabulka je pro rozsáhlost uvedena v příloze č. 4.

#### *Použité metody a výpočty*

Použity byly základní statistické metody absolutní četnosti. Při výběru dat z databáze byl použit základní dotazovací příkaz SELECT s funkcí SUM, ISNULL, GROUP BY,

<sup>37</sup> URL: <[http://cs.artikeldirectory.com/109590\\_Server-Load-Balancing-Take-off-Load/](http://cs.artikeldirectory.com/109590_Server-Load-Balancing-Take-off-Load/)> [cit. 2012-03-05]

ORDER BY. K porovnání dat je použit pruhový spojitý graf, ve kterém byly hodnoty prvního serveru vynásobeny -1, a tím vychýleny pod nulovou osu. Každý pruh spojitého řádku představuje absolutní četnost výskytu dotazů za dobu jednoho měsíce. Modrý pruh představuje počet zpracovaný na Server1 a červený pruh představuje počet zpracovaný na Server2. K posouzení závislosti byl použit výpočet Korelačního koeficientu pomocí funkce CORREL()<sup>38</sup> tabulkového editoru MS EXCEL.



Obrázek 19 - Graf rozložení zatížení pomocí BalancerServeru

### *Forma uplatnění v praxi*

Plná funkčnost BalancerServeru je jedním z hlavních pilířů Cizineckého informačního systému. Zjištění jeho funkčnosti a poměrného rozdělení dotazů na jednotlivé aplikační servery patří mezi nejrelevantnější informace o provozu celého systému. Výstup využijí

<sup>38</sup> URL: < <http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx> > [cit. 2011-12-12]

především pracovníci IT podpory z PČR, kteří vyhodnocují další požadavky směrem k autorské firmě.

Samotný požadavek na BalancerServer je poměrné rozložení zatíženosti obou aplikačních serverů. Interpretací zkonstruovaného grafu je zjištěno, že absolutní vyrovnaní zátěže (počtu dotazů) serverů je uskutečněno v prosinci 2010. Ostatní měsíce vykazují nesouměrnost kolem vertikální osy grafu.

K vypovídající hodnotě závislosti Serveru1 a Serveru2 byl zpracován výpočet **korelačního koeficientu**, jehož hodnota je **-0,77362**. Výpočtem je zjištěna nepřímá úměra a silná závislost mezi jednotlivými zatíženími Serveru1 se Serverem2. Při optimální funkčnosti BalancerServeru je hodnota korelačního koeficientu 1, jelikož počet dotazů obou aplikačních serverů se má rovnat. Další rozbor příčin tohoto jevu není předmětem této práce.

**Vypovídající hodnota** porovnání počtu dotazů v měsíčním intervalu po dobu dvou let na jednotlivé aplikační servery je relevantní údaj, který není ovlivněn objektivní chybou zpracování.

### **4.3.3 Závislost počtu odpovědí na době zpracování dotazu**

Doba zpracování dotazu je hlavním měřítkem pro posuzování funkčnosti informačního systému. Jedním z důvodů navyšování času zpracování dotazu je i počet vrácených odpovědí v příloze, což není nic jiného, než okruh údajů k cizinci podle parametrů výběru. Výstupem tohoto výběru dat budou informace o závislosti počtu odpovědí na čase zpracování dotazu.

#### ***Výběr dat***

Výběr dat je založen na dotazu příkazu SELECT, který vybírá výskyt dotazu na serveru, přičemž sčítá výskyt intervalů doby odezvy (<6 a >5) a zároveň výskyt počtu odpovědí.

Dotaz používá metodu seskupování funkcí GROUP BY a sčítá jednotlivé odchylky pomocí propojení funkce SUM(ISNULL.... .

### ***Dotaz SQL do tabulky tLog***

Pro rozsáhlost dotazu je dotaz SQL do tabulky tLog uveden v příloze číslo 4.

**Tabulka 13 - porovnání četností výskytu odpovědí a délky zpracování dotazu**

	SUMA	Dotaz_T>5	Odpoved_T>5	Relativ
<b>KontrolScot Dotaz</b>	<b>52 758 643</b>	<b>1 806 041</b>	<b>481 118 800</b>	
VýskytCelkem	52 758 643			
P=0	7 483 542			6,72%
P=0_T>5	503 019	503 019		
P=0_T>5_Scot	503 019		503 019	
P1-10	42 253 156			2,04%
P1-10_T>5	861 552	861 552		
P1-10_T>5_Scot	3 007 379		3 007 379	
P11-20	1 147 369			12,90%
P11-20_T>5	147 980	147 980		
P11-20_T>5_Scot	2 167 256		2 167 256	
P21-50	969 852			13,67%
P21-50_T>5	132 536	132 536		
P21-50_T>5_Scot	4 493 531		4 493 531	
P51-100	509 760			8,93%
P51-100_T>5	45 520	45 520		
P51-100_T>5_Scot	3 286 696		3 286 696	
P101-1k	367 322			25,71%
P101-1k_T>5	94 424	94 424		
P101-1k_T>5_Scot	27 034 022		27 034 022	
P1k-10k	22 492			70,72%
P1k-10k_T>5	15 907	15 907		
P1k-10k_T>5_Scot	46 677 572		46 677 572	
P>10k	5 150			99,09%
P>10k_T>5	5 103	5 103		
P>10k_T>5_Scot	393 949 325		393 949 325	

Výsledkem dotazu je tabulka, ve kterém autor porovnává intervaly počtu odpovědí s časem odpovědí, a to nejdříve v absolutních četnostech, a následovně v posledním sloupci v relativních četnostech každého intervalu. První skupina, která představuje žádné odpovědi, má absolutní četnost 503 019 z celkových 7 483 542. V relativním zastoupení se jedná o 6,72% dotazů v kategorii nevyhovující, jelikož doba zpracování je delší než 5 vteřin.

Z další skupiny, která představuje rozsah počtu od 1 do 10 odpovědí na jeden dotaz, je patrné, že z tohoto intervalu je pouze 2,04% dotazů zpracováno déle jak 5 vteřin. Nepřímá úměrnost zjištěných údajů je dobře patrna na následujícím grafu (obrázek číslo 20). Graf znázorňuje modrou spojnici četnost dotazů v závislosti na intervalu četnosti odpovědí. Červená spojnice představuje počet absolutní odpovědí v závislosti na intervalu četnosti odpovědí

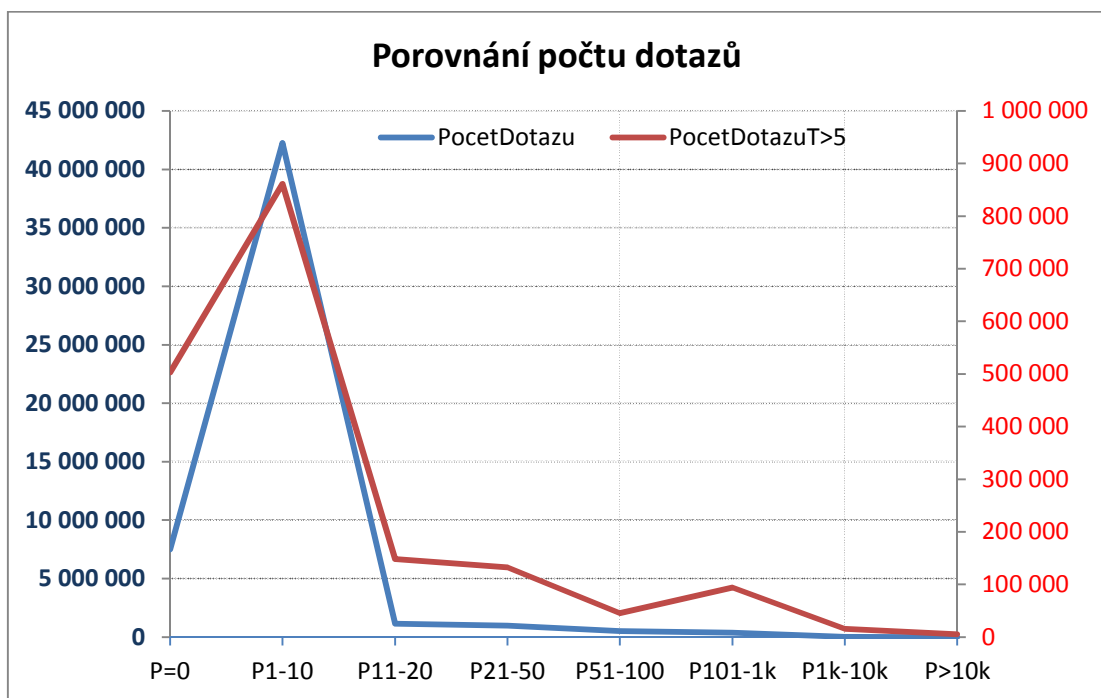


Obrázek 20 – Graf porovnání četnosti dotazů v závislosti na intervalu počtu dotazů

V rozmezí od jedné do deseti odpovědí na jeden dotaz je drtivá většina uskutečněných dotazů. V intervalu nad 10 tisíc odpovědí je minimum dotazů, ale převážná většina

odpovědí, které jsou zpracovávány déle než 5 vteřin. Z příložené tabulky je patrné, že celkem v 99,09% případů jsou dotazy, které vrátí více jak 10 tisíc odpovědí, zpracovávány delší dobu než 5 vteřin. Nepřímo úměrná závislost je tímto potvrzena.

Na obrázku číslo 21 je znázorněno grafické porovnání počtu dotazů v limitu do 5 vteřin a počtu dotazů, které nejsou v limitu do 5 vteřin. Vzhledem k diametrálnímu rozdílu počtů, jsou použity dvě stupnice na vertikální ose grafu. Obě hodnoty jsou porovnávány v závislosti na intervalu počtu odpovědí jednoho dotazu.



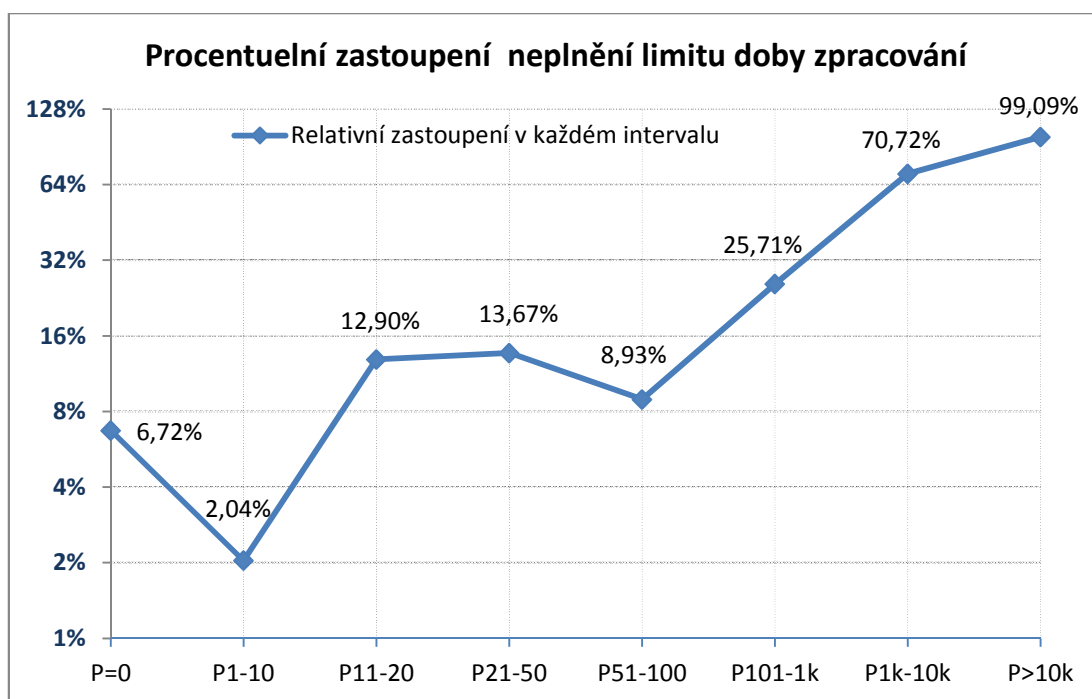
Obrázek 21 - Graf porovnání počtu dotazů zpracovaného v časovém limitu a nad časový limit

### ***Použité metody a výpočty***

Použity byly základní statistické metody absolutní četnosti, relativní četnosti a podobně. Při výběru dat z databáze byl použit základní dotazovací příkaz SELECT s funkcí SUM, ISNULL, GROUPE BY, ORDER BY. Ke grafickému porovnání dat je použit spojnicový graf, v kterém byly použity dvě vertikální osy s odlišným měřítkem. Tímto způsobem bylo dosaženo porovnání výstupů s jiným diametrálním rozlišením.

### Forma uplatnění v praxi

Časové odezvy informačního systému lze ovlivnit výběrem optimalizovaného dotazu. V uvedeném výstupu byl vypočten rozsah četnosti odpovědí, v kterém je ještě splněn limit pěti vteřin. Dotazy, které nesplňují limit zpracování do pěti vteřin, respektive jejich pravděpodobnost nesplnění limitu je vysoká, je potřeba minimalizovat a přístupová práva k takovýmto dotazům zakázat v době největšího zatížení systému. Jejich konkrétní hodnoty pravděpodobnosti v procentech, že v tomto intervalu dojde k nesplnění limitu k zpracování dotazu, jsou znázorněny grafem na obrázku číslo 22. K lepšímu zobrazení nízkých hodnot je použito logaritmického měřítka vertikální osy se základem 2.



Obrázek 22 - Graf procentuálního zastoupení neplnění limitu doby zpracování

**Vypovídající hodnota** porovnání počtu dotazů s počtem odpovědí v závislosti na době potřebné ke zpracování dotazu, je relevantní údaj, který není ovlivněn objektivní chybou zpracování. Ve vybraných datech se ovšem vyskytují údaje o periodickém sehrávání záloh databázového serveru, které čítají řádově miliony odpovědí na jeden dotaz. Ke korektnímu výstupu je nutné tyto údaje vyselektovat. Autor výběru dat a výpočtu jednotlivých

informací nedisponuje popisem struktury logu, který tyto rozsáhlé výběry provádí, proto pro účely diplomové práce nemohly být vyselektovány. V případě rutinního nasazení v praxi je nutné autorskou firmu programu požádat o dešifrování logu hromadného sehrávání a patřičné logy odebrat ze základního souboru.



# 5 Zhodnocení a doporučení

Příprava zdrojů pro diplomovou práci spočívala především v nekonečném úsilí autora získat přístup k části dat bezpečnostního logovacího souboru. Přesvědčování věcného gestora Cizineckého informačního systému o vydání části dat za dobu dvou let provozu, která neobsahuje osobní údaje, bylo nejobtížnějším úkolem autora diplomové práce. Do hloubky přesvědčení o nedotknutelnosti bezpečnostního LogSouboru, přesvědčené kompetentní vedoucí pracovníky, se je podařilo přimět k sepsání podmínek, při jejichž dodržení je možno pro účely diplomové práce data vydat proti podpisu. Po odfiltrování osobních údajů z LogSouboru došlo začátkem roku 2011 k předání dat.

Ihned po předání dat byla provedena transformace z textové podoby do podoby relační databáze. Transformace byla realizována až na několikátý pokus, jelikož při chodu příslušného skriptu docházelo ke kolizím. Kolize spočívaly především v nekorektním datovém typu položky, a používání znaku „@“, který databázový systém používá jako počáteční znak proměnné. V několika ze 48 souborů byly nalezeny chyby v ukončení řádku a nekorektně zapsaného řádku, respektive počtu povinných oddělovačů. Po překonání těchto překážek došlo k odladění databázové tabulky, její indexaci častěji používaných položek a přípravě k další fázi.

Jednotlivé kroky popisují skripty použité při transformaci, výběru dat a výpočtu kontingenčních tabulek za účelem vytvoření konečného výstupu. Na základě grafického znázornění nebo uvedením jednotlivých hodnot jsou informace okomentované významem a doporučené k dalšímu využití v rutinní činnosti jednotlivých složek cizinecké policie.

## 5.1 Nasazení do praxe

Jednotlivé výstupy jsou částečně v rámci Cizinecké policie využívány ve zkušebním provozu pro účely Odboru cizinecké policie při Krajském ředitelství policie hl. m. Praha. V současné době jsou výstupy publikovány v rámci intranetové sítě Ministerstva vnitra pro určený okruh řídicích pracovníků.

K celoplošnému využití v podmínkách Policie České republiky, ale i v jiném státním resortu, je nutné vytvořit interní normativ, jenž přesným způsobem vymezí podmínky, za kterých je možno výstupy zpracovat, distribuovat a stanovit okruh oprávněných osob, které se mohou s jednotlivými informacemi seznamovat. Vypracováním takového normativu se zamezí především dohadům o oprávněnosti získání relevantních dat k dalšímu zpracování.

Další krok postupu spočívá v stanovení zájmových informací pro jednotlivé pracovníky a rozhodnutí o druhu publikování. V rámci Policie České republiky je vybudována vnitřní intranetová síť, v rámci které jsou publikovány jednotlivé webové portály celorepublikových útvarů, webové stránky jednotlivých krajských a okresních útvarů. Právě toto prostředí lze bez jakýchkoliv dalších nákladů využít.

V diplomové práci jsou publikovány veškeré postupy s funkčními skripty, kterými lze data zpracovat a publikovat v tabulkové nebo grafické podobě.

Plnou automatizaci lze realizovat v případě, že distributor dat z CIS bude v pravidelných intervalech ukládat data z LogSouboru na přístupné místo v síti intranet. Automatizace posléze spočívá v namapování primárního zdroje dat a načasování intervalu zpracování jednotlivých publikovaných skriptů.

## **5.2 Ekonomická zátěž**

Po celou dobu realizace jednotlivých postupů, metod, výpočtů a výstupů byl brán prvořadý ohled k ekonomické šetrnosti. Jednotlivé skripty jsou navrhovány v prostředí, které je plně dostupné všem uživatelům výpočetní techniky Policie České republiky. Na úrovni okresních útvarů a jejich nadřízených složek až po útvary Ministerstva vnitra jsou využívány databázové systémy na platformě Microsoft. Přidáním tabulky, jejíž skript pro vytvoření je součástí diplomové práce, je zabezpečena podpora plně automatizovaných propojení s aplikačními servery jednotlivých webových nástrojů. Skripty pro výběr a zpracování dat jsou rovněž publikovány v této práci.

Jakékoliv rozšíření o další výběr informací z LogSouborů lze realizovat pomocí běžných nástrojů, kterými jednotliví pracovníci podpory informačního systému plně disponují.

Na základě uvedených skutečností je zřejmé, že zdroje pro nasazení v praxi jsou k dispozici a není třeba vynakládat žádné finanční prostředky. Vynaložené síly, tudíž práce některých pracovníků při nasazení a údržbě modulů, lze začlenit do každodenních povinností odpovědných osob. Časová náročnost je odhadována na dobu dvou hodin v průběhu jednoho měsíce. V případě plné automatizace zpracování a kontroly s výstražným systémem časová náročnost pracovníka IT odpadá úplně.

### **5.3 Doporučení**

Prvořadým úkolem, při plošném nasazení uvedených modulů, je zajistit jejich právní podložení, vytvořit normativ k vymezení podmínek a stanovení šíře poskytnutých dat a z nich zpracovaných informací. V normativu je nutné zakotvit povinnost věcného gestora ukládat příslušná data na centrální místo a vymežit okruh pracovníků, kteří mohou s daty dále nakládat.

V rámci distribuce dat doporučuje autor v pravidelných intervalech jednoho dne zpracovávat tabulku hodinové četnosti dotazu na serverech, která je publikována v kapitole 4.1.2 i se zdrojem dotazu do SQL. Tuto tabulku kumulativně doplňovat denně o data. Z uvedené tabulky lze při využití dostupných tabulkových editorů získat informace o vytížení evidenčního systému v jednotlivých časových zájmových obdobích, a to od jednohodinového intervalu až po intervaly roční. Tento datový výstup mohou využít při svých statistických výstupech jednotlivé skupiny analytiků k porovnání zatíženosti Cizineckého informačního systému s dalšími statistickými výstupy v rámci jednotlivých útvarů. Velice zajímavé jsou především relevantní informace o vytíženosti pracovní doby v rámci jednoho dne, týdne a měsíce. Jednotlivé výkyvy je možno posléze analyzovat, porovnat s ostatními informacemi a dohledávat příčiny nesouměrnosti pracovního nasazení.

Diplomová práce publikuje jen ukázky možností využití informací ze zpracovaných dat. Další možné rozšíření o větší počet modulů je na každém analytickém pracovníkovi, který mechanismy zpracování ovládá. V případě nutné pomoci, při nasazení do praxe nebo nutnosti změn, je autor ochoten spolupráce.

## 6 Závěr

Cílem diplomové práce bylo vytvoření kanálu zdroje relevantních informací o funkčnosti Cizineckého informačního systému, které by potvrdzovaly nebo vyvracely tvrzení autora a udržovatele programu o jeho spolehlivosti, plné funkčnosti, rychlosti a některých statistických údajích. Relevantní kanál byl touto prací nalezen, vyhodnocen, data transformována do korektní a životaschopné formy. V jednotlivých kapitolách jsou popsány tři až čtyři možnosti využití informací k řídicím, analytickým a diagnostickým účelům s konkrétními příklady výběru dat, výpočty a využití v praxi.

Metody práce byly čerpány především z výsledků bakalářského a magisterského studia na České zemědělské univerzitě v Praze. Dalším zdrojem vědomostí byla dostupná odborná literatura, relevantní webové internetové stránky a osobní praxe autora ze zaměstnání.

Mezi nejdůležitější principy, při tvorbě a zavádění do praxe jednotlivých modulů, řadí autor především ekonomickou nenáročnost. Nenavyšování finančních nákladů, v již tak napjaté době celosvětové ekonomické krize, je řešeno především využitím stávající informační technologie, která je u Policie České republiky provozována. Návrh řešení se dále opírá o síly a prostředky, kterými jsou zajišťovány současné informační systémy. Po zavedení do praxe s plnou automatizací jednotlivých výběrů a výpočtů, nevyžaduje jakoukoliv údržbu, která by byla nad rámec dosavadní činnosti pracovníků informačních technologií. Samotné zavedení lze zvládnout v rámci současného pracovního vytížení, jelikož se převážně jedná o instalační činnosti, které jsou běžně pracovníky prováděny. V případě intuitivní navigace grafického prostředí intranetových stránek s dostatečným komentářem jednotlivých výstupů není potřeba ani provádět školení za účelem pochopení využití jednotlivých výstupů.

Samotná využitelnost informací, které byly výsledkem zpracování dat LogSouboru, byla prokázána již při tvorbě jednotlivých výstupů autorem diplomové práce. Při některých

ukázkách jednotlivých výstupů bylo přímo reagováno na konkrétní potřeby. V době rozhodování o využití finančních prostředků k modernizaci Cizineckého informačního systému bylo rozhodnuto právě na základě dvou statistických výstupů, jež jsou publikovány v diplomové práci.

V dalších několika případech bylo na základě využití výstupů dosaženo včasného odhalení příčin nekorektního průběhu zpracování dotazu. Případ svévolného nekorektního dotazu do CIS byl rovněž odhalen výstupem, který je popisován v této práci.

Zabezpečení chodu velkých informačních systémů u Ministerstva vnitra a jeho podřízených složek, ke kterým patří i Policie České republiky, je čím dál tím více realizováno formou Outsourcingu, což v praxi znamená přenechání této činnosti odborným firmám. Již delší dobu se tímto způsobem provozuje, částečně nebo úplně, informační systém většiny subjektů státní správy.

Správnost tohoto rozhodnutí je předmětem mnoha diskuzí odborníků na celém světě. Menší ekonomické náklady, které v některých případech nejsou tak jednoznačné, jsou porovnávány především s bezpečností uchovávaných dat, reakcí provozního gestora na změny, přiblížení se realitě a podobně. Hodnocení vhodnosti formy Outsourcingu není ovšem náplní této diplomové práce. Cílem je naopak nalézt zdroj relevantních dat, jež mohou po zpracování poskytnout potřebné informace pro jednotlivé řídicí pracovníky útvarů, kteří využívají takto provozovaný informační systém.

Jako zdroj dat byl autorem nalezen bezpečnostní logovací soubor (LogSoubor). Zpracováním dat z bezpečnostního logovacího souboru byl vytvořen způsob, jak v podobě grafického výstup, předat konkrétním pracovníkům relevantní informace k správnému rozhodování při jejich řídicí činnosti.

V širším pojetí výsledků diplomové práce lze konstatovat, že byl nalezen způsob, kterým lze u Outsourcing financovaných informačních systémů, nejenom u Policie České republiky, ale i v ostatním prostředí státní správy, zjišťovat relevantní informace o parametrech aplikačního i datového prostředí, časových limitů a využitelnosti.

Mezi velice zajímavé výstupy by jistě patřilo zpracování bezpečnostních logů informačního systému České správy sociálního zabezpečení v období ledna a února roku 2012, kdy docházelo ke kritickým výpadkům při vyplácení sociálních dávek.

Právě v takovéto situaci vidí autor hlavní důvod, proč hledat kanál k relevantním informacím, které odborným způsobem a prokazatelně změní některé parametry informačního systému. Aniž by autor tušil, že kauza zpoždění výdeje sociálních dávek České správy sociálního zabezpečení vyjde najevo, snažil se řešit tento problém, jelikož k drobným neshodám mezi výrobcem produktu a uživatelem docházelo odjakživa.

## 7 Seznam použitých zdrojů

1. BOLDYŠ, Petr. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 1 – Citace: metodika a obecná pravidla*. Verze 3.3. c1999-2004, poslední aktualizace 11.11.2004. URL: <  
<http://WWW.boldis.cz/citace/citace1.pdf>>.
2. BOLDYŠ, Petr. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 2 – Modely a příklady citací u jednotlivých typů dokumentů*. Verze 3.0 (2004), poslední aktualizace 11.11.2004. URL: <  
<http://WWW.boldis.cz/citace/citace2.pdf>>.
3. BRUCKNER, T. - VOŘÍŠEK, J. *Outsourcing informačních systémů*. EKOPRESS, 1998, ISBN 80-86119-07-6.
4. HUMPHRIES, M., HAWKINS, M. a kol. *Data warehousing - Principy a praxe*. Computer Press 2001. ISBN 80-722-6560-1
5. HOMER, Alexandr, BRIAN, Francis a kol: *Aktive Server Pages 3.0.*, Computer press. 2000. ISBN 80-8609-747-1.
6. MULLER, Scott. *Osobní počítač*. Computer press. 2001. ISBN 80-7226-470-2.
7. SVATOŠOVÁ, Libuše – KÁBA, Bohumil. *Statistické metody I*. Česká zemědělská univerzita v Praze. 2009. ISBN 978-80-213-1688-0.
8. SVATOŠOVÁ, Libuše – KÁBA, Bohumil. *Statistické metody II*. Česká zemědělská univerzita v Praze. 2008. ISBN 978-80-213-1672-0.
9. TAYLOR, A. G. *SQL for dummies. 5th edition*. Indianapolis Willey Publishing. ISBN 07645-4075-0



10. VOSTROVSKÝ, Václav. *Vytváření databází v Oracle*. Česká zemědělská univerzita v Praze. 2008. ISBN 978-80-213-1191-6.
11. URL: <<http://aplikace.mvcr.cz/archiv2008/sbirka/2000/sb032-00.pdf>> [cit. 2011-12-12]
12. URL: <[http://technet.microsoft.com/en-us/library/dd349798\(v=WS.10\).aspx](http://technet.microsoft.com/en-us/library/dd349798(v=WS.10).aspx)> [cit. 2011-12-12]
13. URL: <<http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx>> [cit. 2011-12-12]
14. URL: <[http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA\\_info45/en\\_US/HTML/guide/c-logs.html](http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA_info45/en_US/HTML/guide/c-logs.html)> [cit. 2011-12-12]
15. URL: <[http://technet.microsoft.com/en-us/library/cc786081\(v=ws.10\).aspx](http://technet.microsoft.com/en-us/library/cc786081(v=ws.10).aspx)> [cit. 2011-12-12]
16. URL: <<http://www.w3.org/Daemon/User/Config/Logging.html#LogFormat>> [cit. 2011-12-12]
17. URL: <<http://aplikace.mvcr.cz/archiv2008/sbirka/1999/sb106-99.pdf>> [cit. 2011-12-12]
18. URL: <[http://msdn.microsoft.com/en-us/library/3ca8tfek\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/3ca8tfek(v=vs.85).aspx)> [cit. 2011-12-12]
19. URL: <<http://technet.microsoft.com/en-us/windowsserver/bb512919>> [cit. 2011-12-12]
20. URL: <<http://technet.microsoft.com/en-us/sqlserver/bb671245>> [cit. 2011-12-12]
21. URL: <<http://tools.ietf.org/html/rfc791>> [cit. 2011-12-12]
22. URL: <<http://www.policie.cz/clanek/statistiky-885151.aspx>> [cit. 2011-12-12]

23. URL:< <http://tools.ietf.org/html/rfc791>> [cit. 2011-12-12]
24. URL: < <http://office.microsoft.com/cs-cz/excel-help/correl-HP005209023.aspx>>  
[cit. 2011-12-12]
25. URL: [http://cs.artikeldirectory.com/109590\\_Server-Load-Balancing-Take-off-Load/](http://cs.artikeldirectory.com/109590_Server-Load-Balancing-Take-off-Load/) [cit. 2012-03-05]
26. Policie České republiky. *Závazný pokyn policejního prezidenta č. 159 ze dne 29. prosince 2004, kterým se upravuje postup při provozování informačního systému Policie České republiky služby cizinecké a pohraniční policie*
27. EU. *Smlouva o Evropské unii a smlouvy o fungování Evropské unie v konsolidovaném znění ze dne 15.dubna 2008*
28. *Schengenská prováděcí úmluva v konsolidovaném znění ze dne 5. dubna 2010*

# 8 Přílohy

## 8.1 Příloha 1

### Pomocné tabulky a grafy z kapitoly 4.2.2.

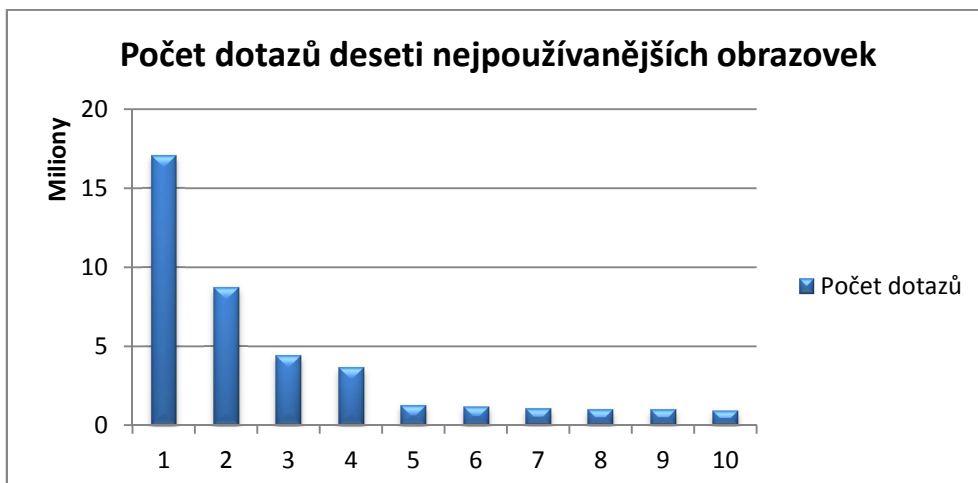
Tabulka 14 - deseti nepoužívanějších obrazovek v CIS a jejich četnost

č.	Název obrazovky	Počet dotazů v roce 2009 - 2010	Počet/730
1	Detail osoby	17059664	23 369
2	Vyhledání	8703365	11 922
3	Podrobná lustrace	4382979	6 004
4	page.label.lus.basic.form	3672010	5 030
5	Základní lustrace osob	1267241	1 736
6	Založení záznamu o ubytování	1194731	1 637
7	Oprava osoby a pobytu	1023373	1 402
8	Vyhledání záznamu víza pro zápis hraniční průvodky	977689	1 339
9	Založení hraniční průvodky	963471	1 320
10	Základní lustrace osob - Výsledek hledání	915071	1 254

Hodnoty v tabulce č. 14 byly vytvořeny SQL dotazem:

```
use simi
select cLogObr AS 'Obrazovka', COUNT(*) AS CELKEM
from tLog
group by cLogObr
```

Po vytvoření základní tabulky byl dopočítán poslední sloupec, kde celkový počet výskytu obrazovky byl vydělen počtem dní za dva roky (730), což jsou hodnoty za jeden den.



Obrázek 23 - graf četnosti deseti nejpoužívanějších obrazovek v CIS

Tabulka 15 - četnost dotazů dle druhu

Druh	Počet
D (výmaz záznamu)	62093
I (přidání záznamu)	5987435
Q (vyhledání záznamu)	43289441
U (oprava záznamu)	3568818

*Hodnoty tabulky č. 15 byly získány dotazem SQL:*

```
use simi
```

```
select cLogDruhD, COUNT(*) AS pocet from tLog
```

```
group by cLogDruhD
```

```
order by cLogDruhD
```

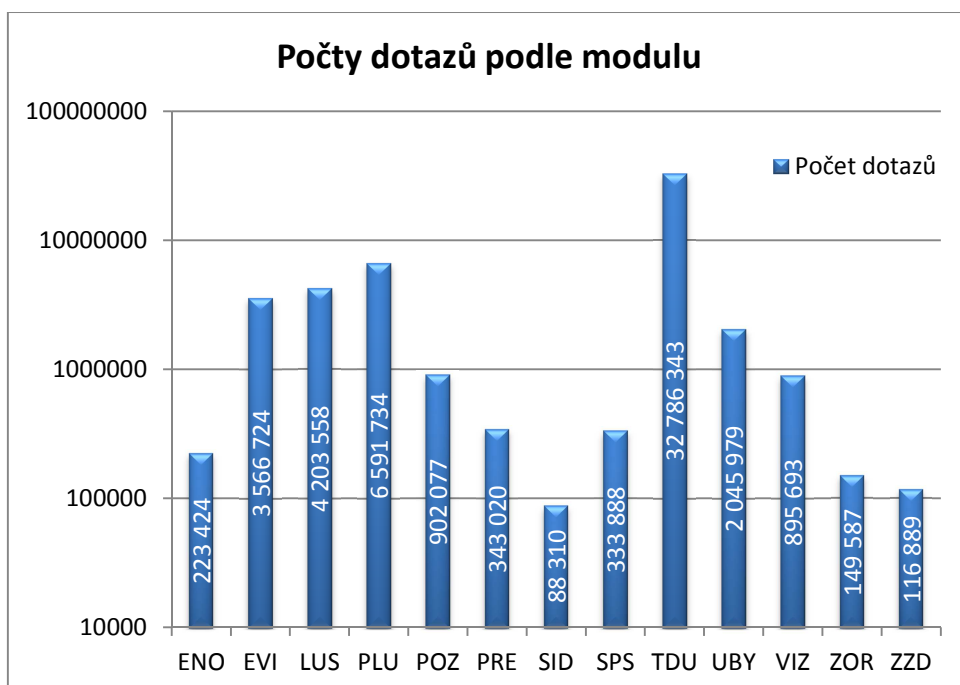
## 8.2 Příloha 2

### *Dotaz SQL pro výběr dat k podkapitole 4.2.3*

```
use simi
SELECT nLogDelkaZprac, COUNT(*) as celkem,
sum(ISNULL(case when nLogDelkaZprac < 3 AND bLogServer=0 then 1 end, 0))
'A_0-2',
sum(ISNULL(case when nLogDelkaZprac < 3 AND bLogServer=1 then 1 end, 0))
'B_0-2',
sum(ISNULL(case when nLogDelkaZprac > 2 AND nLogDelkaZprac < 6 AND
bLogServer=0 then 1 end, 0)) 'A_3-5',
sum(ISNULL(case when nLogDelkaZprac > 2 AND nLogDelkaZprac < 6 AND
bLogServer=1 then 1 end, 0)) 'B_3-5',
sum(ISNULL(case when nLogDelkaZprac > 5 AND nLogDelkaZprac < 11 AND
bLogServer=0 then 1 end, 0)) 'A_6-10',
sum(ISNULL(case when nLogDelkaZprac > 5 AND nLogDelkaZprac < 11 AND
bLogServer=1 then 1 end, 0)) 'B_6-10',
sum(ISNULL(case when nLogDelkaZprac > 10 AND nLogDelkaZprac < 21 AND
bLogServer=0 then 1 end, 0)) 'A_11-20',
sum(ISNULL(case when nLogDelkaZprac > 10 AND nLogDelkaZprac < 21 AND
bLogServer=1 then 1 end, 0)) 'B_11-20',
sum(ISNULL(case when nLogDelkaZprac > 20 AND nLogDelkaZprac < 61 AND
bLogServer=0 then 1 end, 0)) 'A_21-60',
sum(ISNULL(case when nLogDelkaZprac > 20 AND nLogDelkaZprac < 61 AND
bLogServer=1 then 1 end, 0)) 'B_21-60',
sum(ISNULL(case when nLogDelkaZprac > 60 AND nLogDelkaZprac < 121 AND
bLogServer=0 then 1 end, 0)) 'A_61-120',
sum(ISNULL(case when nLogDelkaZprac > 60 AND nLogDelkaZprac < 121 AND
bLogServer=1 then 1 end, 0)) 'B_61-120',
sum(ISNULL(case when nLogDelkaZprac > 120 AND nLogDelkaZprac < 181 AND
bLogServer=0 then 1 end, 0)) 'A_121-180',
sum(ISNULL(case when nLogDelkaZprac > 120 AND nLogDelkaZprac < 181 AND
bLogServer=1 then 1 end, 0)) 'B_121-180',
sum(ISNULL(case when nLogDelkaZprac > 181 AND bLogServer=0 then 1 end,
0)) 'A_>180',
sum(ISNULL(case when nLogDelkaZprac > 181 AND bLogServer=1 then 1 end,
0)) 'B_>180'
from tLog
GROUP BY nLogDelkaZprac
```

Tabulka 16 -Počet dotazů dle délky zpracování

Kum	92,55%	96,56	97,94	98,89	99,63	99,88	99,94	100,00
%	92,55%	4,01%	1,38%	0,94%	0,74%	0,25%	0,06%	0,06%
Počet dotazů	48356557	2093038	723081	493491	388232	131870	31963	28994
Počet dotazů	0s - 2s	3s - 5s	6s - 10s	11s - 20s	21-60	60s - 120s	120s - 180s	> 180s
ENO	218634	2632	1003	478	325	115	47	190
EVI	3492039	25629	13200	11042	22922	302	75	1515
LUS	4185762	14292	1846	891	576	135	7	49
PLU	4786109	827121	362294	276509	212493	87431	18557	21220
POZ	878599	17197	3306	1558	1068	206	54	89
PRE	329848	8554	2661	1190	488	222	29	28
SID	88310	0	0	0	0	0	0	0
SPS	332564	790	291	137	50	24	6	26
TDU	30866760	1181273	334385	199060	146852	41448	12465	4100
UBY	2025154	9946	3010	1946	2949	1713	602	659
VIZ	890679	2316	613	477	313	165	80	1050
ZOR	148076	1266	132	63	44	2	0	4
ZZD	114023	2022	340	140	152	107	41	64



Obrázek 24 - Graf porovnání počtu dotazů podle modulu

## 8.3 Příloha 3

### *Dotaz pro výběr dat z tabulky tLog k podkapitole 4.2.3*

```
use simi
SELECT nLogPocet, COUNT(*) as celkemDotazu,
COUNT(nLogPocet) AS CelkovyPocetOdpovedi,

sum(ISNULL(case when nLogPocet = 0 then 1 end, 0)) 'P_0',
sum(ISNULL(case when nLogPocet = 0 AND nLogDelkaZprac > 5 then 1 end, 0))
'P=0_T>5',

sum(ISNULL(case when nLogPocet = 0 AND nLogDelkaZprac > 5 then nLogPocet end,
0)) 'P=0_T>5_Scot',

sum(ISNULL(case when nLogPocet > 0 AND nLogPocet < 11 then 1 end, 0)) 'P1-10',

sum(ISNULL(case when nLogPocet > 0 AND nLogPocet < 11 AND nLogDelkaZprac > 5
then 1 end, 0)) 'P1-10_T>5',

sum(ISNULL(case when nLogPocet > 0 AND nLogPocet < 11 AND nLogDelkaZprac > 5
then nLogPocet end, 0)) 'P1-10_T>5_Scot',

sum(ISNULL(case when nLogPocet > 10 AND nLogPocet < 21 then 1 end, 0)) 'P11-20',

sum(ISNULL(case when nLogPocet > 10 AND nLogPocet < 21 AND nLogDelkaZprac >
5 then 1 end, 0)) 'P11-20_T>5',

sum(ISNULL(case when nLogPocet > 10 AND nLogPocet < 21 AND nLogDelkaZprac >
5 then nLogPocet end, 0)) 'P11-20_T>5_Scot',

sum(ISNULL(case when nLogPocet > 20 AND nLogPocet < 51 then 1 end, 0)) 'P21-50',

sum(ISNULL(case when nLogPocet > 20 AND nLogPocet < 51 AND nLogDelkaZprac >
5 then 1 end, 0)) 'P21-50_T>5',

sum(ISNULL(case when nLogPocet > 20 AND nLogPocet < 51 AND nLogDelkaZprac >
5 then nLogPocet end, 0)) 'P21-50_T>5_Scot',

sum(ISNULL(case when nLogPocet > 50 AND nLogPocet < 101 then 1 end, 0)) 'P51-100',

sum(ISNULL(case when nLogPocet > 50 AND nLogPocet < 101 AND nLogDelkaZprac >
5 then 1 end, 0)) 'P51-100_T>5',

sum(ISNULL(case when nLogPocet > 50 AND nLogPocet < 101 AND nLogDelkaZprac >
5 then nLogPocet end, 0)) 'P51-100_T>5_Scot',
```

```

sum(ISNULL(case when nLogPocet > 100 AND nLogPocet < 1001 then 1 end, 0)) 'P101-1k',

sum(ISNULL(case when nLogPocet > 100 AND nLogPocet < 1001 AND
nLogDelkaZprac > 5 then 1 end, 0)) 'P101-1k_T>5',

sum(ISNULL(case when nLogPocet > 100 AND nLogPocet < 1001 AND
nLogDelkaZprac > 5 then nLogPocet end, 0)) 'P101-1k_T>5_Scot',

sum(ISNULL(case when nLogPocet > 1000 AND nLogPocet < 10001 then 1 end, 0))
'P1k-10k',

sum(ISNULL(case when nLogPocet > 1000 AND nLogPocet < 10001 AND
nLogDelkaZprac > 5 then 1 end, 0)) 'P1k-10k_T>5',

sum(ISNULL(case when nLogPocet > 1000 AND nLogPocet < 10001 AND
nLogDelkaZprac > 5 then nLogPocet end, 0)) 'P1k-10k_T>5_Scot',

sum(ISNULL(case when nLogPocet > 10000 then 1 end, 0)) 'P>10k',

sum(ISNULL(case when nLogPocet > 10000 AND nLogDelkaZprac > 5 then 1 end, 0))
'P>10k_T>5',

sum(ISNULL(case when nLogPocet > 10000 AND nLogDelkaZprac > 5 then nLogPocet
end, 0)) 'P>10k_T>5_Scot'

from tLog
group by nLogPocet
order by nLogPocet

```



## 8.4 Příloha 4

### *Tabulka porovnání dle zatížení aplikačních serverů – funkcionality BalancerServeru*

Tabulka 17 - porovnání dle rozdělení BalancerServeru.

Mesíc	Celkem	Server 1 (počet přístupů)	Server 2 (počet přístupů)	Pomocne
2009 Leden	2064361	-1285875	778486	1.2009
2009 Únor	1954265	-1172671	781594	2.2009
2009 Březen	2218068	-1338496	879572	3.2009
2009 Duben	2182671	-1333495	849176	4.2009
2009 Květen	2038238	-1320193	718045	5.2009
2009 Červen	2241904	-1372913	868991	6.2009
2009 Červenec	2020101	-1214529	805572	7.2009
2009 Srpen	2099231	-1412907	686324	8.2009
2009 Září	2091352	-873729	1217623	9.2009
2009 Říjen	2090820	-637727	1453093	10.2009
2009 Listopad	3048918	-1973302	1075616	11.2009
2009 Prosinec	1903480	-1308719	594761	12.2009
2010 Leden	1927047	-1158485	768562	1.2010
2010 Únor	2015575	-1086473	929102	2.2010
2010 Březen	2450323	-1156083	1294240	3.2010
2010 Duben	2213016	-735304	1477712	4.2010
2010 Květen	2397192	-1135259	1261933	5.2010
2010 Červen	2359237	-1433864	925373	6.2010
2010 Červenec	2139424	-648422	1491002	7.2010
2010 Srpen	2462601	-448922	2013679	8.2010
2010 Září	2365965	-979609	1386356	9.2010
2010 Říjen	2182393	-1872168	310225	10.2010
2010 Listopad	2307882	-1433599	874283	11.2010
2010 Prosinec	1984569	-960678	1023891	12.2010

## 8.5 Příloha číslo 5

Zákon České republiky č. 326/1999 Sb. O pobytu cizinců na území České republiky ve znění jeho posledních změn. Zákonný rozsah informačního systému cizinecké policie. Právní podpora Cizineckého informačního systému.

### HLAVA XV

#### INFORMAČNÍ SYSTÉMY

##### § 158

(1) Policie při výkonu působnosti podle tohoto zákona provozuje informační systém cizinců, jehož je správcem a který obsahuje jméno, popřípadě jména, příjmení, včetně dřívějších jmen a příjmení, den, měsíc a rok narození, pohlaví a státní příslušnost cizince, jeho rodné číslo, nebo jiný identifikační údaj sdělený tímto cizincem, obrazový záznam, například fotografii, cizince a jeho daktyloskopické otisky. Dále se v informačním systému cizinců o cizinci vedou

a) údaje v rozsahu

1. žádosti o udělení víza ([§ 54](#)), povolení k dlouhodobému pobytu ([§ 42 odst. 5](#)), povolení k trvalému pobytu ([§ 70 odst. 1](#)),
2. žádosti o vydání potvrzení o přechodném pobytu na území, pobytové karty rodinného příslušníka občana Evropské unie, průkazu o povolení k trvalému pobytu občana Evropské unie nebo průkazu o povolení k trvalému pobytu ([§ 87x](#)),
3. žádosti o vydání cizineckého pasu, cestovního průkazu totožnosti a cestovního dokladu podle [§ 108 odst. 1 písm. f](#)) ([§ 111 odst. 1](#)),
4. výjezdního příkazu ([§ 50 odst. 6](#)), hraniční průvodky ([§ 14](#)), tiskopisu pozvání ([§ 180 odst. 2](#)) a přihlašovacího tiskopisu ([§ 97](#)),

b) údaje o

1. dni, měsíce a roku vydání víza nebo dokladu podle tohoto zákona, jeho číslu, druhu a době platnosti,
2. dni, měsíce a roku ověření pozvání nebo jeho odepření,
3. dni, měsíce a roku, kdy nabylo právní moci rozhodnutí o zamítnutí žádosti o vydání dokladu podle tohoto zákona a o důvodu zamítnutí,
4. dni, měsíce a roku vydání výjezdního příkazu, jeho číslu a o době jeho platnosti,
5. zrušení, zániku nebo skončení platnosti víza nebo dokladu vydaného podle tohoto zákona, včetně uvedení jeho druhu a číslu, a důvodu zrušení, zániku nebo skončení platnosti,
6. hodině, dni, měsíce, roku a místu překročení státních hranic a o odepření vstupu nebo odepření vycestování,
7. důvodu zařazení do evidence nežádoucích osob s uvedením doby platnosti omezení vstupu na území,
8. důvodu povolení vstupu na území podle [§ 122 odst. 1](#) nebo [2](#),
9. zařazení do informačního systému smluvních států jiným smluvním státem s uvedením doby platnosti omezení vstupu na území smluvních států a o orgánu jiného státu, který

cizince do systému zařadil,

10. dni, měsíce a roku, kdy bylo vydáno rozhodnutí o správním nebo soudním vyhoštění, a kdy toto rozhodnutí nabylo právní moci, a o důvodu pro jeho vydání,

11. dni, měsíce a roku překročení státních hranic při průvozu přes území a o orgánu, který o průvoz požádal, a orgánu, který provedení průvozu umožnil,

12. dni, měsíce a roku nabytí právní moci rozhodnutí o správním deliktu podle tohoto zákona, označení správního deliktu podle zákonného ustanovení, skutkových okolnostech případu podle výroku rozhodnutí, způsobu vyřízení a údaje vztahující se k vybírání a vymáhání uložené pokuty,

13. druhu, číslu a době platnosti ztraceného nebo odcizeného dokladu a o datu ohlášení ztráty nebo odcizení dokladu,

c) údaje

1. umožňující určit orgán, který rozhodoval ve věci, o které se vedou údaje podle písmene b),

2. umožňující určit spis vedený k cizinci a založený ve věci, o které se vedou údaje podle písmene b),

3. vztahující se k průběhu řízení ve věci, o které se vedou údaje podle písmene b), a výkonu rozhodnutí podle tohoto zákona,

d) údaje o

1. zbavení nebo omezení způsobilosti k právním úkonům,

2. rodinném stavu, datu a místě uzavření manželství, datu nabytí právní moci rozhodnutí soudu o prohlášení manželství za neplatné, datu nabytí právní moci rozhodnutí soudu o neexistenci manželství, datu zániku manželství smrtí jednoho z manželů, nebo datu nabytí právní moci rozhodnutí soudu o prohlášení jednoho z manželů za mrtvého a dni, který byl v pravomocném rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který manžel prohlášený za mrtvého nepřežil, anebo datu nabytí právní moci rozhodnutí soudu o rozvodu manželství,

3. datu a místě vzniku partnerství, datu nabytí právní moci rozhodnutí soudu o neplatnosti nebo o neexistenci partnerství, datu zániku partnerství smrtí jednoho z partnerů, nebo datu nabytí právní moci rozhodnutí soudu o prohlášení jednoho z partnerů za mrtvého a dni, který byl v pravomocném rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který partner prohlášený za mrtvého nepřežil, anebo datu nabytí právní moci rozhodnutí soudu o zrušení partnerství,

4. datu, místu a okrese úmrtí; jde-li o úmrtí mimo území, o státu, ve kterém k úmrtí došlo, popřípadě datu úmrtí,

5. dni, který byl v rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který cizinec prohlášený za mrtvého nepřežil,

jde-li o cizince s povoleným trvalým pobytem na území, s povoleným přechodným pobytem na území na dobu delší než 90 dnů, o občana Evropské unie, který na území hodlá přechodně pobývat po dobu delší než 3 měsíce, a o cizince, kterému byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>,

e) údaje o

1. stupni osvojení,

2. původním a novým jménem, popřípadě jménech, a příjmení dítěte,

3. původním a novým rodným čísle dítěte,

4. datu a místa narození,
5. rodných číslech osvojitelů; v případě, že osvojiteli nebylo přiděleno rodné číslo, údaje o jménu, popřípadě jménech, příjmení a datu narození osvojitele,
6. rodných číslech otce a matky; pokud jim nebylo přiděleno, údaje o jejich jménu, popřípadě jménech, příjmení a datu narození; tyto údaje se nevedou, pokud se jedná o dítě narozené ženě s trvalým pobytem na území, která porodila dítě a písemně požádala o utajení své osoby v souvislosti s porodem,
7. datu nabytí právní moci rozhodnutí o osvojení nebo rozhodnutí o zrušení osvojení dítěte, jde-li o cizince s povoleným trvalým pobytem na území, s povoleným přechodným pobytem na území na dobu delší než 90 dnů, o občana Evropské unie, který na území hodlá přechodně pobývat po dobu delší než 3 měsíce, a o cizince, kterému byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>,

f) údaje o jménu, popřípadě jménech, a příjmení

1. zletilého cizince, který je nezaopatřeným dítětem cizince,
2. nezletilého cizince, který byl cizinci s oprávněním k pobytu na území nebo jeho manželu rozhodnutím příslušného orgánu svěřen do náhradní rodinné péče, nebo který byl cizincem s oprávněním k pobytu na území nebo jeho manželem osvojen anebo jehož poručníkem nebo manželem jeho poručníka je cizinec,
3. osamělého cizince staršího 65 let nebo bez ohledu na věk cizince, který se o sebe nedokáže ze zdravotních důvodů sám postarat, jde-li o sloučení rodiny s rodičem nebo dítětem s oprávněním k pobytu na území,
4. cizince, který je nezaopatřeným přímým příbuzným ve vzestupné nebo sestupné linii nebo takovým příbuzným manžela občana Evropské unie,
5. rodiče nezletilého cizince, kterému byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>, a jeho rodné číslo; jde-li o cizince, kterému nebylo přiděleno rodné číslo, jméno, popřípadě jména, příjmení a datum narození,

g) biometrické údaje zpracované v nosiči dat průkazu o povolení k pobytu,

h) záznam o poskytnutí údajů.

(2) Informační systém cizinců podle [odstavce 1](#) obsahuje rovněž údaje o cizincích, kterým byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>, a to v rozsahu údajů uvedených v [odstavci 9](#); tyto údaje poskytuje do informačního systému cizinců ministerstvo.

(3) Na žádost cizince lze v informačním systému cizinců podle [odstavce 1](#) vést údaj o adrese, na kterou mu mají být doručovány písemnosti podle tohoto zákona nebo zvláštního právního předpisu. Adresu podle věty první, její změnu, nebo zrušení, může cizinec ohlásit policii nebo ministerstvu.

(4) Policie při výkonu působnosti podle tohoto zákona dále provozuje informační systémy obsahující údaje o

a) dni, měsíce a roku nabytí právní moci rozhodnutí o správním deliktu podle tohoto zákona, jehož pachatelem je právnická nebo podnikající fyzická osoba, označení správního deliktu podle zákonného ustanovení, skutkových okolnostech případu uvedených

ve výroku rozhodnutí a údaje vztahující se k vybírání a vymáhání uložené pokuty,

b) zvoucí osobě v rozsahu tiskopisu pozvání ([§ 180 odst. 2](#)), jménu, příjmení, dni, měsíce a roku narození a státní příslušnosti zvaného cizince, o skutečnostech, které byly důvodem pro odepření ověření pozvání a o porušení závazku přijatého v pozvání, a údaje o orgánu, který ve věci jednal,

c) číslu a druhu dokladu, který byl orgánem vydávajícího státu prohlášen za neplatný, včetně důvodu neplatnosti.

(5) Policie je oprávněna provozovat další informační systém, pokud je jeho provozování nezbytnou podmínkou plnění úkolu podle tohoto zákona. V tomto informačním systému mohou být obsaženy toliko údaje získané při plnění tohoto úkolu.

(6) Údaje vedené v informačních systémech podle [odstavců 4](#) a [5](#) mohou být obsaženy i v informačním systému cizinců podle [odstavce 1](#), pokud mají vztah k cizinci, který je v tomto informačním systému evidován.

(7) Údaj z informačního systému vedeného podle tohoto zákona lze sdružovat s údajem obsaženým v jiném informačním systému vedeném podle zvláštního právního předpisu, pokud je to nezbytné pro plnění úkolu uloženého tímto zákonem.

(8) Ministerstvo je oprávněno v rámci své působnosti stanovené tímto zákonem vkládat do informačního systému cizinců podle [odstavce 1](#) a [odstavce 4 písm. c\)](#) údaje získané v rámci plnění úkolů ministerstva a údaje z informačních systémů podle [odstavců 1, 4](#) a [5](#) využívat pro svou činnost.

(9) Z informačního systému cizinců podle [odstavce 1](#) se poskytují tyto údaje:

- a) jméno, popřípadě jména, příjmení, rodné příjmení,
- b) datum narození,
- c) pohlaví,
- d) místo a stát, kde se cizinec narodil; v případě, že se narodil na území, místo a okres narození,
- e) rodné číslo,
- f) státní občanství, popřípadě státní příslušnost,
- g) druh a adresa místa pobytu na území, popřípadě adresa, na kterou mají být doručovány písemnosti podle jiného právního předpisu,
- h) číslo a platnost oprávnění k pobytu,
- i) počátek pobytu, popřípadě datum ukončení pobytu na území,

- j) zbavení nebo omezení způsobilosti k právním úkonům,
- k) správní nebo soudní vyhoštění a doba, po kterou není cizinci umožněn vstup na území,
- l) zařazení do státního integračního programu, jde-li o cizince s udělenou mezinárodní ochranou,
- m) rodinný stav, datum a místo uzavření manželství, datum nabytí právní moci rozhodnutí soudu o prohlášení manželství za neplatné, datum nabytí právní moci rozhodnutí soudu o neexistenci manželství, datum zániku manželství smrtí jednoho z manželů, nebo datum nabytí právní moci rozhodnutí soudu o prohlášení jednoho z manželů za mrtvého a den, který byl v pravomocném rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který manžel prohlášený za mrtvého nepřežil, anebo datum nabytí právní moci rozhodnutí soudu o rozvodu manželství,
- n) datum a místo vzniku partnerství, datum nabytí právní moci rozhodnutí soudu o neplatnosti nebo o neexistenci partnerství, datum zániku partnerství smrtí jednoho z partnerů, nebo datum nabytí právní moci rozhodnutí soudu o prohlášení jednoho z partnerů za mrtvého a den, který byl v pravomocném rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který partner prohlášený za mrtvého nepřežil, anebo datum nabytí právní moci rozhodnutí soudu o zrušení partnerství,
- o) jméno, popřípadě jména, příjmení, státní občanství, popřípadě státní příslušnost manžela nebo partnera a jeho rodné číslo; je-li manžel nebo partner cizinec, který nemá přiděleno rodné číslo, jméno, popřípadě jména, příjmení a datum narození,
- p) jméno, popřípadě jména, příjmení, státní občanství, popřípadě státní příslušnost dítěte, pokud je cizincem, a jeho rodné číslo; v případě, že dítěti nebylo rodné číslo přiděleno, jméno, popřípadě jména, příjmení a datum narození,
- q) jméno, popřípadě jména, příjmení, státní občanství, popřípadě státní příslušnost otce, matky, popřípadě jiného zákonného zástupce, pokud jsou cizinci, a jejich rodné číslo; v případě, že jeden z rodičů nebo jiný zákonný zástupce nemá přiděleno rodné číslo, jméno, popřípadě jména, příjmení a datum narození,
- r) o osvojeném dítěti, pokud je cizincem:
1. stupeň osvojení,
  2. původní a nové jméno, popřípadě jména, příjmení dítěte,
  3. původní a nové rodné číslo dítěte,
  4. datum a místo narození, státní občanství, popřípadě státní příslušnost,
  5. rodná čísla osvojitelů; v případě, že osvojiteli nebylo přiděleno rodné číslo, údaje o jménu, popřípadě jménech, příjmení a datu narození osvojitele,
  6. rodná čísla otce a matky; pokud jim nebylo přiděleno, údaje o jejich jménu, popřípadě jménech, příjmení a datu narození; tyto údaje se nevedou, pokud se jedná o dítě narozené ženě s trvalým pobytem na území, která porodila dítě a písemně požádala o utajení své osoby v souvislosti s porodem,

7. datum nabytí právní moci rozhodnutí o osvojení nebo rozhodnutí o zrušení osvojení dítěte,

s) datum, místo a okres úmrtí; jde-li o úmrtí mimo území, stát, ve kterém k úmrtí došlo, popřípadě datum úmrtí,

t) den, který byl v rozhodnutí soudu o prohlášení za mrtvého uveden jako den smrti, popřípadě jako den, který cizinec prohlášený za mrtvého nepřežil,

u) jméno, popřípadě jména, příjmení

1. zletilého cizince, který je nezaopatřeným dítětem cizince,

2. nezletilého cizince, který byl cizinci s oprávněním k pobytu na území nebo jeho manželku rozhodnutím příslušného orgánu svěřen do náhradní rodinné péče, nebo který byl cizincem s oprávněním k pobytu na území nebo jeho manželem osvojen anebo jehož poručníkem nebo manželem jeho poručníka je cizinec,

3. osamělého cizince staršího 65 let nebo bez ohledu na věk cizince, který se o sebe nedokáže ze zdravotních důvodů sám postarat, jde-li o sloučení rodiny s rodičem nebo dítětem s oprávněním k pobytu na území,

4. cizince, který je nezaopatřeným přímým příbuzným ve vzestupné nebo sestupné linii nebo takovým příbuzným manžela občana Evropské unie,

5. rodiče nezletilého cizince, kterému byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>, a jeho rodné číslo; jde-li o cizince, kterému nebylo přiděleno rodné číslo, jméno, popřípadě jména, příjmení a datum narození,

pokud jsou tyto údaje vedeny o cizincích s povoleným trvalým pobytem na území, s povoleným přechodným pobytem na území na dobu delší než 90 dnů, o občanech Evropské unie, kteří na území hodlají přechodně pobývat po dobu delší než 3 měsíce, a o cizincích, kterým byla udělena mezinárodní ochrana nebo dočasná ochrana podle zvláštního právního předpisu<sup>2)</sup>, a to v rozsahu stanoveném zvláštním právním předpisem a způsobem umožňujícím dálkový přístup. Za podmínek stanovených tímto zákonem se poskytuje rovněž záznam o poskytnutí údajů.

(10) V informačním systému cizinců se o údajích uvedených v [odstavci 9](#) zpracovávají a vedou rovněž jejich změny, a to včetně dat, ke kterým tyto změny nastaly, jsou-li policii známa.

(11) O poskytnutí údajů z informačního systému cizinců se provede záznam o datu a hodině výdeje údajů a jejich poskytnutém rozsahu s uvedením označení orgánu veřejné moci, kterému byly poskytnuty.

(12) Z informačního systému cizinců se zahraničním třetím osobám ani orgánům cizích států neposkytují podle [odstavce 9](#) údaje o cizincích s udělenou mezinárodní ochranou.

(13) V případech, kdy byly podle [odstavce 9](#) na základě zvláštního právního předpisu poskytnuty údaje za účelem zajišťování

a) bezpečnosti státu,

- b) obrany,
- c) veřejného pořádku a vnitřní bezpečnosti,
- d) předcházení, vyšetřování, odhalování a stíhání trestných činů,
- e) významného hospodářského nebo finančního zájmu České republiky nebo Evropské unie včetně měnové, rozpočtové a daňové oblasti, nebo

f) ochrany subjektu údajů

a subjekt, kterému byly tyto údaje poskytnuty, písemně prohlásí, že záznam o poskytnutí údajů nelze po dobu trvání skutečností podle písmen a) až f) zpřístupnit, lze tak učinit výlučně pro plnění úkolů orgánu činnému v trestním řízení, jde-li o trestný čin související s účelem poskytnutí údajů, nebo orgánu vykonávajícímu dozor nad zpracováním osobních údajů podle jiného právního předpisu. Policie má k tomuto záznamu o poskytnutí údaje přístup pouze v rozsahu nezbytně nutném pro účely výkonu činnosti správce podle zvláštního právního předpisu.

(14) Subjekt, kterému mají být poskytnuty údaje podle [odstavce 13](#), je povinen policii oznámit údaje umožňující identifikaci subjektu a identifikaci fyzické osoby požadující poskytnutí údajů jeho jménem a dále oznámit, kdy byla ukončena doba trvání zajištění účelu uvedeného v [odstavci 13](#), pro který by mu byly údaje poskytnuty. Prohlášení podle [odstavce 13](#) a oznámení podle věty první mohou být zaslána i elektronickými prostředky.

(15) Subjekty, kterým jsou podle zvláštního právního předpisu poskytovány osobní údaje podle [odstavce 9](#),

- a) nejsou oprávněny k jejich shromažďování, předávání a využívání mimo působnost stanovenou zvláštním právním předpisem a
- b) jsou povinny zajistit ochranu dat před náhodným nebo neoprávněným přístupem nebo zpracováním.

(16) Policie využívá údaje vedené v jiných informačních systémech veřejné správy pro účely zjištění správnosti údajů vedených v informačním systému cizinců.

(17) Policie poskytuje z informačního systému cizinců podle [odstavce 1](#) ministerstvu za účelem správy informačního systému datových schránek údaje v rozsahu

- a) jméno, popřípadě jména, příjmení, případně jejich změna, rodné příjmení,
- b) datum narození,
- c) místo a stát, na jehož území se narodil,
- d) rodné číslo, bylo-li přiděleno, a jeho změny,



- e) státní občanství,
- f) adresa místa pobytu na území,
- g) zbavení nebo omezení způsobilosti k právním úkonům,
- h) datum úmrtí,
- i) den, který byl v rozhodnutí soudu o prohlášení za mrtvého uveden jako den úmrtí.