

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Cloudové řešení v EEG/ERP doméně

Plzeň, 2013

Roman Tryml

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 28. 4. 2013

.....

Poděkování:

Tímto vyjadřuji poděkování vedoucímu práce Ing. Petru Ježkovi, Ph.D. za cenné rady a připomínky, velkou vstřícnost a ochotu mi vždy poradit a pomoci.

Abstract:

The aim of this thesis is to find a suitable cloud solution for applications and databases that are developed by the neuroinformatics research group focusing at the Department of Computer Science and Engineering of the Faculty of Applied Science of the University of West Bohemia. The first part of the thesis explains possibilities of cloud computing. It includes overview and analysis of Amazon, Google and INCF solutions.

The second part describes a particular solution the INCF Dataspace; procedure and experience with its connecting to the incf zone is presented. The implementation includes connecting the data from the EEG/ERP Database of experiments in to the Dataspace system is described.

Obsah

1	Úvod	1
2	Výzkumná skupina neuroinformatiky	2
2.1	Základní informace	2
2.2	Požadavky pro výběr cloudového řešení	2
3	Problematika cloud computingu	4
3.1	Stručný přehled	4
3.2	Vývojové trendy v informačních technologiích	5
4	Analýza vybraných poskytovatelů	6
4.1	Řešení od Amazon	6
4.2	Řešení Google	15
4.3	Řešení Microsoft	18
4.4	Řešení BeeScale	20
4.5	Řešení International Neuroinformatics Coordinating Facility (INCF)	23
5	Zhodnocení analyzovaných poskytovatelů	24
6	INCF Dataspac e a systém iRODS	26
6.1	Základní charakteristika iRODS	26
6.2	Klient systému iRODS	27
6.3	Připojení do systému INCF Dataspac e	29
7	Získání dat z databáze EEG pro INCF Dataspac e	32
7.1	Úvodní informace	32
7.2	Řešení pomocí jazyka PL/SQL	33
7.3	Řešení pomocí jazyka Java	33
7.4	Zhodnocení řešení	34
8	Závěr	36

1 Úvod

Na Katedře informatiky a výpočetní techniky Fakulty aplikovaných věd Západočeské univerzity v Plzni působí několik výzkumných skupin. Mezi nimi i skupina se zaměřením na výzkum neuroinformatiky. Experimenty prováděné touto skupinou produkují velké množství dat, která potřebují být k dispozici nejen výzkumníkům ze ZČU, ale i odborníkům z celého světa. Výzkumná skupina na KIV se rozhodla pro využití jednoho z nových trendů v informačních technologiích, a to cloud computingu (dále také cloud). Cloud může být plnohodnotným alternativním řešením k nákupu vlastního informačního systému. Jednou z možností využití cloud computingu je schopnost zajistit ukládání či sdílení dat. Úkolem této bakalářské práce je provést analýzu možností cloud computingu pro skupinu se zaměřením na výzkum neuroinformatiky KIV.

V následujících kapitolách budou zmíněni někteří poskytovatelé, jejichž nabídky cloudu budou vyzkoušeny a otestovány. Na základě porovnání těchto nabídek bude podle požadavků výzkumné skupiny vybrán nejvhodnější poskytovatel cloudu. Další část se pak zaměří na komplexní popis vybraného řešení a zdůvodnění, proč bylo vybráno. Bude zvolen konkrétní postup instalace a zprovoznění domény EEG¹/ERP² v cloudu. O tomto postupu budou uvedeny reference s uvedením konkrétních úskalí a jejich řešení. Výsledek budou podrobeny praktickému testování s vyhodnocením.

Práce obsahuje celkem tři hlavní témata, přičemž každé se skládá z teoretického vysvětlení problematiky a popisu praktických zkušeností. První téma pojednává o cloud computingu. Budou teoreticky popsány současné možnosti a praktické vyzkoušení některého z řešení. Druhé téma popisuje systém iRODS a Dataspace a postup připojení do zóny incf. Třetí téma pojednává o výběru dat z databáze neuroinformatické skupiny pro sdílení v systému Dataspace. Následuje popis praktické realizace.

¹Electroencephalography - Je vyšetření registrující aktuální elektrickou aktivitu mozku pomocí povrchových elektrod umístěných na povrchu lebky.

²Event-Related Potential – Evokovaný potenciál je mozková odezva na nějakou událost, tzv. stimul.

2 Výzkumná skupina neuroinformatiky

V této kapitole bude uvedena stručná charakteristika neuroinformatické výzkumné skupiny na KIV a dále zde budou popsána její kritéria a požadavky pro výběr optimálního cloudového řešení.

2.1 Základní informace

Výzkumná skupina při Katedře informatiky a výpočetní informatiky Západočeské univerzity v Plzni se specializuje na výzkum medicínských aplikací, zejména na oblast neurologie. S ohledem na tento výzkum jsou hojně využívány metody elektroencefalografie (EEG) a evokované potenciály (ERP). Výzkumná skupina využívá webový EEG/ERP portál jako centrální úložiště pro data/metadata získaná při EEG/ERP experimentech, a software JERPA - desktopovou aplikaci pro výpočetně náročné operace. EEG/ERP experimenty jsou obvykle časově náročné a produkují velké množství dat, proto se musí aktivně řešit jejich dlouhodobé skladování a management. Problémem je praktické sdílení a výměna dat mezi jednotlivými výzkumnými laboratořemi z různých míst světa tak, aby byla zároveň zachována jejich bezpečnost. [1]

2.2 Požadavky pro výběr cloudového řešení

Únosná finanční náročnost řešení

Výzkumná skupina je součástí státní vysoké školy, proto jsou její finanční možnosti velmi omezené. Z tohoto důvodu je vhodné najít poskytovatele, který podporuje vědu a výzkum malou, v ideálním případě nulovou finanční náročností svých cloudových služeb.

Kompatibilita cloudového řešení s EEG/ERP doménou

EEG/ERP portál je pro zajištění vysoké úrovně abstrakce vyvíjen v programovacím jazyce Java i s využitím XML technologie. Pro zpracování velkých datových souborů slouží databázový server Oracle 11g. Proto je nutné najít poskytovatele, který podporuje tyto informační systémy.

Dostupnost služby cloudového řešení

Dalším požadavkem výzkumné skupiny je 100% dostupnost cloudových služeb. Je nutné brát v úvahu taková řešení, která v tomto směru uživatele neomezují například variabilním výkonem svého cloudu v závislosti na jeho aktuálním zatížení. Vhodné je vybírat poskytovatele s ohledem na budoucnost.

Dostačující datové úložiště

Vzhledem k zmíněnému velkému množství informací produkovaných EEG/ERP experimenty je nutné nalézt řešení, které neomezuje velikost ukládaných dat.

Zabezpečení citlivých dat

Jelikož se jedná o osobní data, je nutné vybrat takového poskytovatele, který plně garantuje jejich zabezpečení proti zcizení a zneužití.

Jednoduchost implementace a snadná administrace

Jednoduchost implementace je zde relativní, záleží na zaměření, znalostech a zkušenostech administrátora. Důraz by však měl být kladen na jednoduchost a nenáročnost pro koncové uživatele z řad výzkumníků. Systém by měl být ideálně přístupný odkudkoliv z internetu (se zřetelem na bezpečnostní autentizaci uživatele systému).

3 Problematika cloud computingu

Následující kapitola přinese popis samotného cloudu, jeho členění, využití a stručný přehled některých poskytovatelů.

3.1 Stručný přehled

Cloud computingu je na Internetu založený model vývoje a používání počítačových technologií. Lze jej charakterizovat jako pronájem informačního systému, jako služby od poskytovatelů, přičemž tyto služby je možné rozdělit podle jejich typu do následujících skupin: [2]

SaaS - Software as a Service (software jako služba)

Poskytovatel v tomto modelu pronajímá aplikace včetně infrastruktury (hardware a operačního systému/běžového prostředí). Můžeme využívat připravené aplikace poskytovatelem služby.

PaaS - Platform as a Service (platforma jako služba)

Poskytovatel nabízí běžové prostředí pro provoz a využití vlastních aplikací. Nevýhodou je však proprietální uzavření poskytovatelem: API rozhraní, programovací jazyk, databáze většinou s přesně danou nebo omezenou možností výběru. Tuto nevýhodu řeší poslední model cloudové služby, a to:

IaaS - Infrastructure as a Service (infrastruktura jako servis)

Je zde nabízen výpočetní výkon, úložiště a síťová konektivita - hardware pro naše vlastní běžové prostředí a aplikace.

Výhody cloud computingu jsou zejména:

- vždy aktuální software
- platba za skutečně využívané služby
- škálovatelnost, možnost rychlé změny zdrojů podle konkrétní potřeby
- dostupnost z Internetu a mobilních zařízení

Nevýhody cloud computingu:

- ztráta úplné kontroly nad vlastními daty
- nutnost zabezpečení osobních dat šifrováním
- obtížná nebo finančně nákladná migrace k jinému poskytovateli
- stabilita informačního systému závisí na pronajímateli

Dále se cloud dělí na:

Public Cloud (Veřejný cloud)

Pronájem informačního systému poskytuje externí společnost.

Private Cloud (Soukromý cloud)

Společnost v tomto modelu cloudu provozuje vlastní informační systém jako službu. Management má tomto případě k dispozici přehled využívání informačních služeb dle oddělení, např. pro analytické účely nákladů na IT systém.

Hybridní cloud

Spojení obou výše uvedených typů cloudů. Pokud aktuálně, myšleno nárazově, systémové prostředky vlastního informačního systému společnosti nepostačují, je možné využít veřejného cloudu.

3.2 Vývojové trendy v informačních technologiích

Vývoj trendů v informačních technologiích je velmi rychlý. Informace, které byly před několika měsíci novinkou, nemusí být nyní již aktuální. Stojíme tedy před problémem, jak k této skutečnosti přistupovat. Obecné informace a definice týkající se novinek lze čerpat prostřednictvím knižní literatury. Pro hlubší analýzu je potřeba pročitat aktuální články či se účastnit tematických konferencí apod.

Cloud computing je typickým příkladem rychle se vyvíjejícího moderního trendu v informačních technologiích. Informace potřebné pro popis cloudových řešení uvedených v této práci byly čerpány z aktuálních internetových článků, konferencí a osobních konzultací s odborníky ze společností poskytujících cloud computing.

4 Analýza vybraných poskytovatelů

Tato kapitola přináší konkrétní nabídky společností poskytující cloud computing. Pokud to bude možné, konkrétní řešení bude otestováno a zhodnoceno.

4.1 Řešení od Amazon

Společnost Amazon je průkopníkem na trhu cloudových služeb. Začínala s internetovým obchodem. V roce 2002 spustila na volných hardwarových prostředcích službu AWS¹. [3] V roce 2006 byla spuštěna služba on-line úložného prostoru Amazon Simple Storage Service (S3) a představena služba Amazon Elastic Compute Cloud (EC2). [4] AWS představuje kompletní sadu infrastrukturních a aplikačních služeb, jejichž seznam a stručný popis následuje níže.

Výpočetní výkon:

EC2 (Amazon Elastic Compute Cloud)

- nabízí výpočetní kapacitu v Cloudu.

Auto Scaling

- umožňuje automaticky škálovat výpočetní výkon v závislosti na podmínkách, které potřebujeme nebo definujeme.

Elastic Load Balancing

- automaticky distribuuje příchozí aplikační provoz na více Amazon EC2 instancí.

Databáze:

Amazon Relational Database Service (RDS)

- umožňuje snadné nastavení, ovládání a škálování relační databáze v Cloudu.

Amazon DynamoDB

- je plně řízená, vysoce výkonná NoSQL²databáze, kterou lze snadno nastavit, provozovat a škálovat.

Amazon SimpleDB

- představuje řízenou NoSQL databázi určenou pro menší soubory dat.

Nasazení a správa:

¹AWS - Amazon Web Services

²NoSQL - databáze nevyužívající jazyk SQL

AWS Identity a Access Management (IAM)

- IAM umožňuje bezpečně řídit přístup k AWS službám a zdrojům uživatelů, umožňuje vytváření a správu uživatelů v AWS a umožňuje udělit přístup k AWS zdrojům pro uživatele spravované mimo AWS, například v podnikovém adresáři.

Síťové služby:

Amazon Virtual Private Cloud (VPC)

- umožňuje poskytování soukromé izolované části Amazon Web Services (AWS), kde je možné spustit AWS zdroje ve virtuální síti. S Amazon VPC můžeme definovat virtuální topologii sítě, která se podobá té tradiční a která by zároveň mohla pracovat v našem vlastním datovém centru.

Datová úložiště

Amazon Simple Storage Service (S3)

- poskytuje infrastrukturu pro plně redundantní ukládání a načítání libovolného množství dat kdykoli a odkudkoli na webu.

Amazon Glacier

- skladovací služba s extrémně nízkými náklady, která poskytuje bezpečné a trvanlivé úložiště pro archivaci dat a zálohování.

Podpora:

AWS Support

- rychlý kanál pro pomoc se sestavením a spuštěním aplikací v AWS infrastruktuře.

Níže následuje podrobnější popis služby Amazon EC2 (Elastic Compute Cloud) a Simple Storage Service(S3).

Amazon EC2 je webová služba, která nabízí variabilní výpočetní kapacitu. Výhoda tohoto řešení spočívá v tom, že využíváme pouze výpočetní výkon, který skutečně potřebujeme. Přidání nové instance serveru není složité a umožňuje rychle škálovat kapacitu výpočetního výkonu nahoru nebo dolů. Máme možnost vybrat si z několika typů instancí. Standardní instanci první generace (M1) si můžeme představit, jako pronajmutý virtuální server různé "velikosti". Příklad malého virtuálního serveru: 1,7 GiB paměti, 1 EC2 Compute Unit (1 virtuální jádro s 1 EC2 výpočetní jednotkou), 160 GB místní instance úložiště, 32-bit nebo 64-bit platformy. Příklad velkého virtuálního serveru: 7,5 GiB paměti, 4 EC2 výpočetního jednotky (2 virtuální jádra každá s 2 EC2 výpočetními jednotkami), 850 GB místní instance úložiště, 64-bit platformy.

Pokud by standardní virtuální servery první generace byly nedostačující, existuje možnost vybrat si z generace druhé (M3). Tedy z ještě výkonějších virtuálních serverů. Příklad dvojité extra velké instance serveru z druhé generace je 30 GiB paměti, 26 EC2 výpočetních jednotek (8 virtuálních jader s 3,25 EC2 výpočetními jednotkami), EBS úložiště, 64-bit platformy. Všechny virtuální servery jsou samozřejmě k dispozici ihned, platí se pouze za jejich využití. To je podstatný rozdíl oproti vlastnímu podnikovému serveru.

Pokud však nechceme měnit celé virtuální servery, ale chceme pouze škálovat výkon po menších výpočetních jednotkách, může zcela dobře posloužit Micro instance. Micro instance (t1.micro) poskytuje malé množství konzistentních prostředků procesorů a umožňují zvýšit kapacitu CPU v krátkých dávkách. Jsou vhodné pro nižší propustnosti aplikací a webových stránek, které pravidelně vyžadují delší výpočetní cykly. Problémem není ani aplikace s velkými nároky na paměť nebo na velký výkon procesoru. Vše je totiž možné libovolně podřídit konkrétním požadavkům. Pokud je jeden z požadavků na rendering nebo mediální zpracování aplikace, lze pro tyto účely využít grafické procesory (GPU) s masivním paralelním zpracováním. Využitelné je například pro grafické studio při tvorbě animací či celých animovaných filmů bez potřeby vlastního hardware. Požadujeme-li extrémně rychlý přístup k datům (vysoce výkonné databázové úlohy), jsou k dispozici SSD disky s vysokou I/O propustností.

Amazon S3 poskytuje jednoduché rozhraní webových služeb, které lze použít ukládáním a načítáním libovolného množství dat, a to kdykoliv a odkudkoliv na webu. Velikost objektů se může pohybovat od 1B až do 5TB, přičemž počet objektů je neomezený. Výhodou pro optimalizaci latence je možnost zvolit si umístění úložiště dle regionu. Je garantováno, že například data uložená v Irsku (pro region EU), nikdy tuto oblast neopustí. Nezbytností jsou autentizační mechanismy s šifrováním dat. Vlastník těchto dat obdrží určitý „klíč“, bez kterého se k datům nikdo jiný nedostane.

Ceník služeb Amazon

Služby Amazon jsou bohužel zpoplatněny. Proto považuji za důležité analyzovat i ceník, protože právě cena může být kritériem při výběru cloudového poskytovatele. Před samotnou finanční analýzou je dobré znát alespoň přibližný interval náročnosti na výpočetní výkon používaných aplikací, velikost a rychlost úložiště dat, síťový provoz (čímž je myšleno i časové využití aplikace a zatížení uživateli), požadavky na software, respektive operační systém a běhové prostředí pro aplikaci. Pro vyzkoušení běhu doporučuji vyzkoušení provozu v cloudu od

společnosti Amazon službu AWS Free na rok zdarma. [6] Pro celkově lepší přehlednost cen jsou zpracovány tabulky níže.

Balíček AWS Free obsahuje:

- 750 hodin měsíčně provozu Amazon EC2 (Linux nebo Windows) s 613 Mb paměti
- 750 hodin měsíčně provozu Load Balanceru s 15 Gb úložného prostoru
- 30 Gb Block Storage s 2 000 000 I/O a 1 Gb prostoru pro snapshot
- 750 hodin Amazon RDS, relační databáze (MySQL, Oracle BYOL¹ nebo SQL Server) s úložným prostorem o velikosti 20 Gb a 10 000 000 I/O operací.
- další služby jako 5 Gb Simple Storage Service (S3), Dynamo DB, Simple Workflow (SWF), Simple Queue Service (SQS) a Simple Notification Service (SNS), CloudWatch, přenos dat a Data Pipeline
- kromě těchto služeb je k dispozici též zdarma i AWS Management Console, která pomůže vytvořit a spravovat aplikace na AWS

Tyto volné vrstvy jsou k dispozici pouze pro nové zákazníky, a to po dobu 12 měsíců po přihlášení k účtu AWS Free. Pokud volné použití vyprší a/nebo pokud běžící aplikace přesahuje volné použití, je nutno zaplatit standardní pay-as-you-go službu dle ceníku.

Ceny služeb jsou rozděleny do několika kategorií, např. podle regionu, který je pro daný cloud vybrán, podle délky závazku využití služeb nebo pokud postačí cloud s dynamicky se měnící volnou kapacitou prostředků Amazonu, kterou mají aktuálně k dispozici.

On-Demand instance [4] umožňuje platit za výpočetní kapacity na hodinu bez paušální platby. Není nutné plánovat výpočetní kapacitu s dlouhodobým výhledem, nakupovat a udržovat hardware.

¹BYOL - Bring Your Own License

	Region EU (Irsko)	
Standard On-Demand Instances	Linux/UNIX	Windows
Small (Default)	\$0.085 za hodinu	\$0.115 za hodinu
Medium	\$0.170 za hodinu	\$0.230 za hodinu
Large	\$0.340 za hodinu	\$0.460 za hodinu
Extra Large	\$0.680 za hodinu	\$0.920 za hodinu

Tab.1 Ceník On-Demand instance

V tabulce Tab.2 je uveden ceník dalších typů On-Demand instancí pro aplikace se speciálními nároky na paměť, výkon procesoru, GPU nebo rychlejší I/O přenosy.

	Region EU (Irsko)	
Micro On-Demand Instances	Linux/UNIX	Windows
Micro	\$0.020 za hodinu	\$0.035 za hodinu
High-Memory On-Demand Instances		
Extra Large	\$0.506 za hodinu	\$0.570 za hodinu
Double Extra Large	\$1.012 za hodinu	\$1.140 za hodinu
High-CPU On-Demand Instances		
Medium	\$0.186 za hodinu	\$0.285 za hodinu
Extra Large	\$0.744 za hodinu	\$1.140 za hodinu
Cluster GPU Instances		
Quadruple Extra Large	\$2.36 za hodinu	\$2.60 za hodinu
High-I/O On-Demand Instances		
Quadruple Extra Large	\$3.410 za hodinu	\$3.580 za hodinu

Tab.2 Ceník On-Demand instancí pro speciální požadavky

Další možností volby jsou vyhrazené instance (**Reserved Instances**) virtuálních serverů v cloudu. Pokud bychom věděli, že danou instanci budeme s určitostí využívat delší dobu, např. jeden rok, zaplatíme jednorázovou částku a poté podobně jako u On-Demand instancí hodinovou sazbu, která je ale v tomto případě výrazně levnější. Vyhrazené instance se dále dělí na tři skupiny dle výše jednorázové platby na Light, Medium a Heavy. Ceníky vyhrazených instancí pro porovnání jsou uvedeny v tabulkách Tab. 3, Tab.4, Tab.5, Tab.6, Tab.7 a Tab.8. Všechny ceny jsou uvedeny pro Linux. Pro Windows jsou paušální ceny stejné. Hodinová sazba pro Windows je však o \$0.2 - \$0.4 vyšší než pro Linux.

LIGHT Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Standart Reserved Instances				
Small (Default)	\$69	\$0.049	\$106.30	\$0.039
Medium	\$138	\$0.098	\$212.50	\$0.078
Large	\$276	\$0.196	\$425.20	\$0.156
Extra Large	\$552	\$0.392	\$850.40	\$0.312

Tab.3 Ceník vyhrazených standartních instancí Light pro Linux

LIGHT Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Micro Reserved Instances				
Micro	\$23	\$0.015	\$35	\$0.015
High-Memory Reserved Instances				
Extra Large	\$353	\$0.288	\$548	\$0.23
Double Extra Large	\$706	\$0.576	\$1096	\$0.46
Cluster GPU Reserved Instances				
Quadruple Extra Large	N/A*	N/A*	N/A*	N/A*
High-I/O Reserved Instances				
Quadruple Extra Large	\$2576	\$2.605	\$3884	\$2.035

* Není k dispozici v regionu EU (Irsko) - nutné zvolit jiný region

Tab.4 Ceník vyhrazených instancí Light pro Linux pro speciální požadavky aplikací

MEDIUM Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Standart Reserved Instances				
Small (Default)	\$160	\$0.031	\$250	\$0.025
Medium	\$320	\$0.063	\$500	\$0.05
Large	\$640	\$0.124	\$1000	\$0.10
Extra Large	\$1280	\$0.248	\$2000	\$0.20

Tab.5 Ceník vyhrazených standartních instancí Light pro Linux

MEDIUM Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Micro Reserved Instances				
Micro	\$54	\$0.01	\$82	\$0.01
High-Memory Reserved Instances				
Extra Large	\$850	\$0.185	\$1283	\$0.148
Double Extra Large	\$1700	\$0.37	\$2566	\$0.296
Cluster GPU Reserved Instances				
Quadruple Extra Large	N/A*	N/A*	N/A*	N/A*
High-I/O Reserved Instances				
Quadruple Extra Large	\$5973	\$1.379	\$9133	\$1.022

* Není k dispozici v regionu EU (Irsko) - nutné zvolit jiný region

Tab.6 Ceník vyhrazených instancí Light pro Linux pro speciální požadavky aplikací

HEAVY Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Standart Reserved Instances				
Small (Default)	\$195	\$0.025	\$300	\$0.02
Medium	\$390	\$0.05	\$600	\$0.04
Large	\$780	\$0.10	\$1200	\$0.8
Extra Large	\$1560	\$0.20	\$2400	\$0.16

Tab.7 Ceník vyhrazených standartních instancí Light pro Linux

HEAVY Reserved Instances	Region EU (Irsko)			
	na 1 rok		na 3 roky	
	předem	za hodinu	předem	za hodinu
Micro Reserved Instances				
Micro	\$62	\$0.008	\$100	\$0.008
High-Memory Reserved Instances				
Extra Large	\$1030	\$0.148	\$1550	\$0.118
Double Extra Large	\$2060	\$0.296	\$3100	\$0.236
Cluster GPU Reserved Instances				
Quadruple Extra Large	N/A*	N/A*	N/A*	N/A*
High-I/O Reserved Instances				
Quadruple Extra Large	\$7280	\$0.931	\$10960	\$0.742

* Není k dispozici v regionu EU (Irsko) - nutné zvolit jiný region

Tab.8 Ceník vyhrazených instancí Light pro Linux pro speciální požadavky aplikací

Platba předem za vyhrazenou instanci je navratná. Po dobu, kdy nebude instance Light nebo Medium spuštěna, se nehradí hodinové sazby. Pouze u Heavy instance je účtována hodinová sazba bez ohledu na její spuštění. Pokud se společnosti Microsoft (Windows) nebo RedHat (Linux) rozhodnou zvýšit licenční poplatky za software, vyhrazuje si společnost Amazon úměrně tomuto nárůstu navýšit hodinové sazby. Je zde pak samozřejmě možnost odstoupit od využívání služby Amazon a poměrná část platby bude vrácena.

Poslední instancí, kterou je vhodné zmínit, je **Spot Instance**. Spot instance umožňuje využít aktuální volnou kapacitu Amazon EC2. Instance je zpoplatněna spotovou cenou, která je stanovená Amazon EC2 a kolísá periodicky v závislosti na nabídce a poptávce po kapacitě Spot instance. Pokud bychom chtěli využít Spot instanci, zašleme dotaz s uvedením typu instance na dostupnost požadované zóny včetně uvedení počtu Spot instancí, které chceme spustit, a maximální cenou, kterou jsme ochotni zaplatit za hodinu. Tato maximální cena se porovnává s aktuálními spotovými cenami. Historie Spot cen je také k dispozici na Amazon EC2 API a AWS Admin Console. Pokud maximální cena přesahuje aktuální spotovou cenu, je žádost splněna a naše instance poběží, dokud se ji rozhodneme ukončit, nebo spotová cena přesáhne námi určený maximální finanční limit.

Pokud bychom chtěli z přehledových tabulek porovnat finanční nákladnost jednotlivých instancí, zvolíme si například On-Demand Large instanci. On-Demand je bez platby předem. Hodinová sazba v tomto případě vychází na \$0.460. Kurz dolaru předpokládejme 20 Kč. Hodinová sazba v Korunách českých vychází na 9,20 Kč. Denní sazba činí 220 Kč. Měsíční náklady na instanci serveru od Ama-

zone představují částku 6624 Kč. Jedná se o cloud server se 7,5 GiB paměti, 4 EC2 výpočetní jednotky (2 virtuální jádra každá s 2 EC2 výpočetními jednotkami), 850 GB místní instance úložiště, 64-bit platforma. Pokud bychom oproti tomu počítali s nákupem odpovídajícího serveru, s dobou čekání na dodávku, jeho provozní náklady, spotřeba elektrické energie, údržba a životnost, představuje řešení od Amazon.com zajímavou alternativu.

Pro vyzkoušení služeb je nutné provést registraci na webových stránkách¹. V dalším kroku je jako nezbytnou podmínku nutno uvést údaje o platební kartě, což mnou nebylo akceptováno. Z tohoto důvodu praktické otestování služby AWS Free nebylo možné.

¹<http://www.amazon.com>

4.2 Řešení Google

Google je americká nadnárodní společnost, která od doby svého vzniku v roce 1998 poskytuje internetové produkty a služby. Do podvědomí vstoupila především svým vyhledávačem. Společnost se v současné době zaměřuje mimo jiné i na vývoj software, například prohlížeč Chrome a operační systém Android. Aktuálně se společnost Google snaží proniknout i do oblasti cloud computingu, protože vlastní volné výpočetní kapacity, které nejsou plně využity. Investice do cloud computingu se zde proto jeví jako výhodně řešení využití volných kapacit a zároveň získání určitých finančních prostředků z jejich pronájmu. Vzhledem k tomu, že Google sám vznikl na akademické půdě, snaží se jako společnost poskytovat školství a výzkumu veškeré služby zdarma. Z toho důvodu je v této práci společnost Google zahrnuta do analýzy možných cloudových řešení pro neuroinformatickou doménu. Společnost Google v současnosti poskytuje tyto cloudové možnosti:

Google Apps

- představuje soubor hotových online aplikací (firemní email, kalendáře, rezervace, kontakty a skupiny, chat, telefony, videokonference, webové stránky, disk Google – sdílení v rámci firmy). Aplikace lze rovnou nastavit a ihned používat. Toto řešení patří do modelu cloud computingu SaaS (Software as a Service). Nevýhodou toho řešení je nemožnost použití vlastního software, lze využívat pouze předdefinovaný. [7] Tento problém řeší následující produkt:

Google App Engine (GAE)

- nabízí prostor pro běh vlastních aplikací a software. Tento model cloudové služby spadá do kategorie PaaS (Platform as a Service). I toto řešení má však svá úskalí. Aplikace musí být napsaná v jazycích určených poskytovatelem, tj.: Java, Python a experimentálně Go (vlastní programovací jazyk společnosti Google). GAE dále nenabízí klasické instalační metody jako je FTP nebo webové rozhraní. Aplikace je nutné nahrát a spustit pomocí speciálních nástrojů. GAE celkově představuje velmi dobré řešení pro běh aplikací Java a Python, musíme však vzít v úvahu možnou nevýhodu v podobě Googlem omezených, respektive ne vždy stejně velkých výpočetních a datových kapacit. Dostupný výpočetní výkon je nabízen v závislosti na aktuální zátěži serverů společnosti Google, nemáme jej tedy pro naši aplikaci vždy garantovaný, nebo dostupný. [8]

Google Compute Engine (GCE)

- je poslední novinkou společnosti Google v oblasti cloud computing. GCE byl v červnu 2012 představen na konferenci společnosti Google v San Franciscu v Kalifornii. Tento produkt spadající do modelu cloudových služeb IaaS (Infrastructure

as a Service) umožňuje spouštět náročné výpočetní úlohy na virtuálních serverech s operačním systémem Linux, na vlastní infrastruktuře společnosti Google. Za zmínku jistě stojí fakt, že na uvedené konferenci byla při prezentaci GCE spuštěna demo aplikace z oblasti biologie, která v nárazovém maximu dosáhla zátěže šest set tisíc CPU jader. GCE má být přímým konkurentem EC2 společnosti Amazone (viz výše). Vzhledem k datu uvedení tohoto cloudového řešení na trh nejsou zcela dostupné detailní informace o produktu a zkušenosti uživatelů s ním samotným. Pokud by se však potvrdily nastíněné možnosti a GCE opravdu takto zvládal velmi náročné výpočetní operace, dá se reálně předpokládat, že by GCE bylo výhodnější alternativou clonovému řešení společnosti Amazone. [9] Praktické otestování GCE (pro účely této práce) je nemožné z důvodu nutnosti zakoupit předplatné tzv. Google Gold Support - uživatelské podpory.

Google Cloud Storage (GCS)

-představuje virtuální datové úložiště, které je možné využít pro data naší aplikace v celkovém rozsahu až 10 GB. Možnosti této služby jsou velmi široké. Lze ji využít přímo pro data naší aplikace, jako záloha nebo replikace datových serverů v případě jejich výpadku. [10]

Google BigQuery

-je systém, který umožňuje analyzovat uživatelská data v řádech TB bez nutnosti využití vlastního hardware či software. Nahrávání dat je velmi rychlé, protože systém využívá API, které používá souběžně komprimované streamy. [11]

Google Cloud SQL

-reprezentuje databázový systém v cloudu, který umožňuje vytvářet, spravovat a konfigurovat relační databáze MySQL. Umožňuje výběr mezi servery v USA a EU, podporuje programovací jazyky Java a Python. Velikost databáze se pohybuje do 100 GB a paměť pro transakce 16 GB RAM pro každého uživatele. Negativem je absence podpory uživatelských funkcí a MySQL replikace. Společnost tento produkt nabízí k vyzkoušení až do 1. června 2013 zcela zdarma. [12]

Produkty společnosti Google nabízejí velmi zajímavou alternativu cloudového řešení Amazone svými cenovými relacemi, respektive bezplatným využitím pro akademickou sféru. Zmíněná bezplatnost je ověřována pouze z internetových zdrojů. Ač jsem kontaktoval společnost NetMail, která je certifikovaným partnerem Google Apps pro Českou republiku, neobdržel jsem na dotaz odpověď. Nevýhodou cloudového řešení společnosti Google je kromě výše zmíněných nedostatků také fakt, že vzhledem ke konkurenčnímu boji se společností Microsoft Google příležitostně ukončuje podporu některých svých služeb (například ukončení podpory Ex-

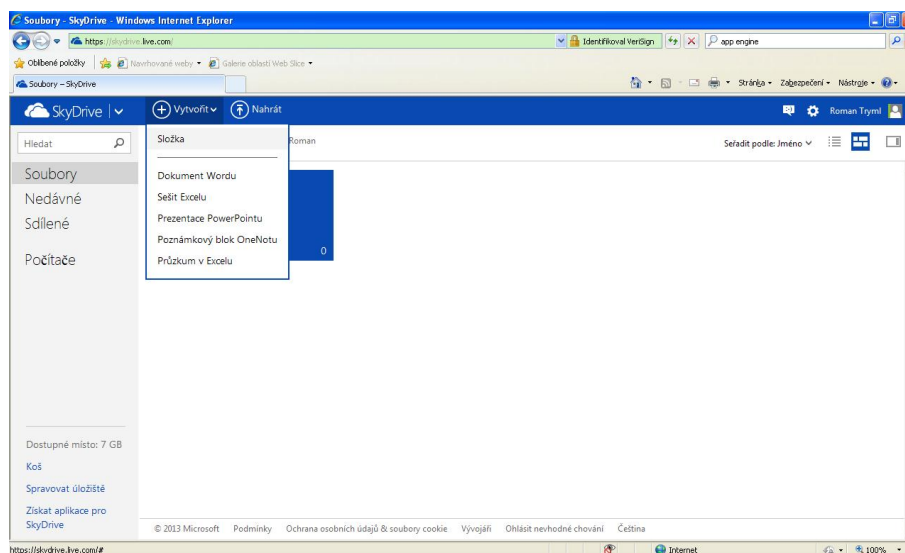
change ActiveSync v prosinci 2012). [13] Problém může i činit určitá netransparentnost smluvních podmínek (například vynucené používání některých ne zcela souvisejících služeb, případně vynucené poskytnutí soukromých dat či informací, které nutně nesouvisí s poskytovanou službou). Skutečnost, že společnost Google nabízí relativně široké spektrum cloudových možností, může být paradoxně nevýhodou, neboť Google zákazníkovi nenabízí ucelený přehled a popis poskytovaných služeb (oproti společnosti Amazone). Zjišťování informací o jednotlivých cloudových službách této společnosti není příliš jednoduché. Z tohoto důvodu jsem své poznatky ověřovat osobní konzultací s Ing. Šedivým, který řadu let pracoval ve společnosti Google.

4.3 Řešení Microsoft

Společnost Microsoft zachytila trend pronájmu informačních technologií později. Vzhledem ke skutečnosti, že se jedná o jednu z největších softwarových firem na světě, je vhodné její řešení uvést. Existuje zde předpoklad nejen smazání deficitu, ale i dalšího rychlého rozvoje nabízených služeb. Hlavní cloudovou platformu společnosti Microsoft představuje Windows Azure. Jako datové úložiště Microsoft poskytuje službu SkyDrive.

Microsoft Azure je možné využít pro hostování webových aplikací. Jedná se o jednoduché webové aplikace i vícevrstvé služby, kdy systém škáluje svůj výkon na základě aktuálního zatížení webu. Microsoft nabízí zdarma 90denní zkušební verzi Windows Azure pro otestování, ovšem podobně jako Amazon požaduje registraci včetně údajů z platební karty. Bez těchto informací není možné přihlášení korektně dokončit. [14]

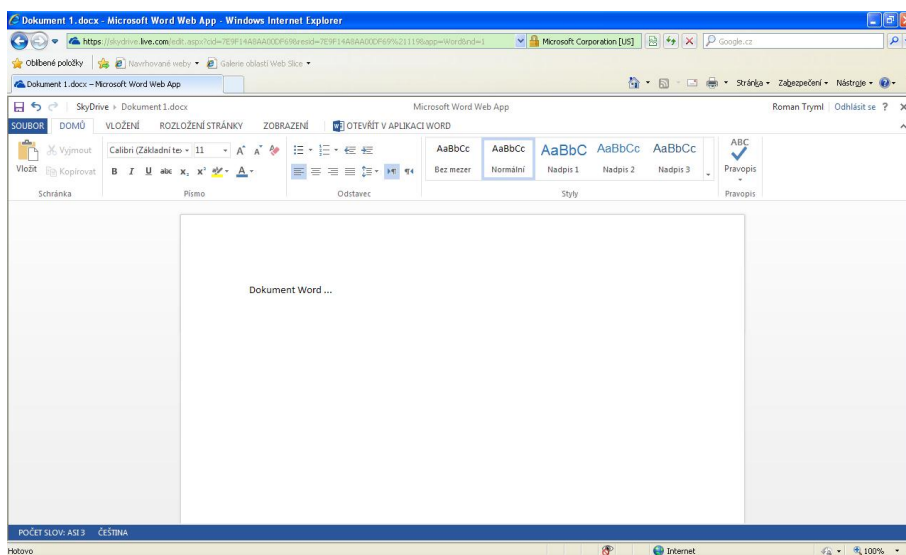
Microsoft nabízí v rámci služby SkyDrive zdarma 7 GB volného datového prostoru. [15] Diskový prostor je dostupný po přihlášení na webových stránkách¹. Při registraci je požadována pouze platná emailová adresa. Součástí služby SkyDrive je možnost vytvářet vlastní dokumenty známe z balíku MS Office. Ukázka webového přístupu na službu SkyDrive je znázorněna na obrázku 1.



Obrázek 1: Webový přístup k službě MS SkyDrive

¹<https://skydrive.live.com>

Připojení k úložišti je možné dvěma způsoby: přes webové rozhraní a pomocí aplikace, kterou je možné nainstalovat do počítače. Zpřístupní se tímto způsobem SkyDrive jako další disk. Osobně jsem vyzkoušel přístup přes webové rozhraní, které je svým vzhledem a grafickým uspořádáním zaměřeno na koncového zákazníka, u nějž se předpokládá, že ke svým datům přistupuje i pomocí dotykových obrazovek tabletů a „chytrých“ telefonů. Obsluhu zvládne i běžný uživatel počítače. Vytváření dokumentů MS Office je také jednoduché. Dokumenty lze ponechat v cloudovém úložišti SkyDrive a je umožněn i jejich přenos do vlastního počítače. Na obrázku č. 2 je zobrazena ukázka vytvoření dokumentu Word ve SkyDrive. Z obrázku je zřejmé identické prostředí jako dektopová aplikace MS Word.

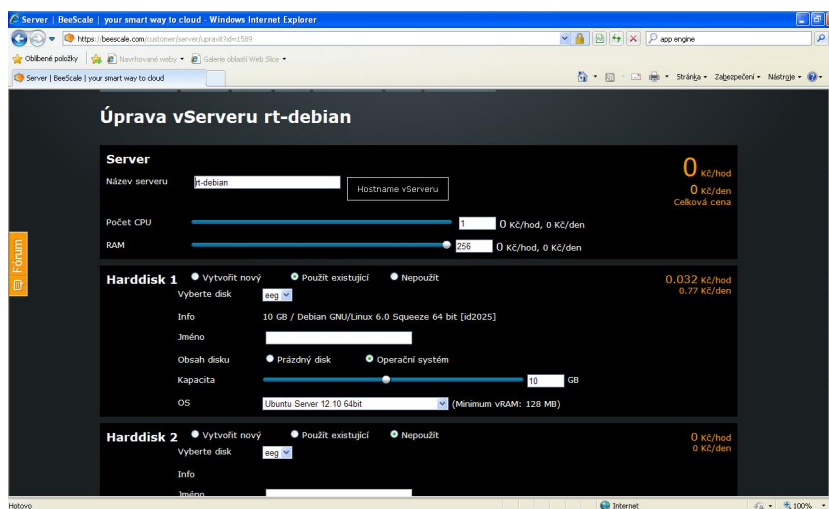


Obrázek 2: Vytvoření dokumentu MS Word ve SkyDrive

4.4 Řešení BeeScale

Jak je zřejmé výše uvedeného, cloudová řešení s výjimkou MS SkyDrive nebylo vždy možné korektně vyzkoušet a prakticky si ověřit uvedené informace. Služba BeeScale společnosti OldanyGroup umožňuje po registraci získat volný kredit na vyzkoušení serveru v cloudu. Registrovat účet je možné na webových stránkách¹. Je možné vytvořit plnohodnotné servery s předinstalovanými operačními systémy Ubuntu, Debian, FreeBSD, CentOS a Turnkey. Připravují se i operační systémy Windows. Je možné pronajmout si libovolný počet serverů, každý s 1 - 4 virtuálními procesory, 64 MB - 4 GB paměti, 1 - 4 pevnými disky, každý o kapacitě 1 až 100 GB. [16]

Postup vytvoření vlastního serveru k cloudu je následující. Na výše uvedené webové stránce provedeme registraci. Je nutné zadat platnou emailovou adresu. Získáme 100 Kč kreditu zdarma. Nyní vytvoříme server, zadáme název serveru, určíme počet CPU a velikost RAM. V rámci zkušební verze máme k dispozici pouze 1 vCPU o rychlosti 3 GHz a maximální velikost vRAM 256 MB. Zadáme počet disků. Kapacita disku ve zkušební verzi je 10GB. Zvolíme přeinstalovaný operační systém. Konfigurace serveru je zřejmá z obrázku č.3.



Obrázek 3: Konfigurace virtuálního serveru služby BeeScale

Parametry serveru jsou nastaveny a je jej tedy možné spustit. Spuštění serveru provedeme z dialogu *Serverovna* jak je zřejmé s obrázku č.4.

¹<https://beescale.com/>



Obrázek 4: Dialogové okno vypnutého serveru služby BeeScale

Na obrázku č. 5 je vidět webový dialog spuštěného serveru. Z dialogu je možné vyčíst parametry pro připojení jako je ip adresa serveru, je možné zobrazit heslo pro uživatele root. Server lze restartovat nebo úplně vypnout. Pokud je server vypnutý, nedochází k odečítání kreditu.



Obrázek 5: Dialogové okno spuštěného serveru služby BeeScale

Od tohoto okamžiku je server s operačním systémem Debian plně k dispozici. Pro administraci je přístupný přes ssh protokol. Soubory je možné přenášet přes

sftp. Je možný přístup uživatele root. Konfigurace balíčků aplikací je plně v naší režii. Nainstaloval jsem na server vývojové a běhové prostředí Java a aplikační server Apache Tomcat. Vše je bez problémů funkční.

Cena za služby odpovídá množství pronajatých prostředků, které jsou využívány. Na webové stránce¹ je přehledně zobrazena hodinová taxa za spotřebované služby a aktuální výše kreditu s údajem, za jaké časové období bude vyčerpán (při současném průměrném využití systému). Při plné zátěži testované konfigurace serveru je udaná částka 0,27 Kč/h, denní sazba je tedy 6,35 Kč, měsíční přibližně 197 Kč. Tato částka je maximální. Pokud není server 100% využit, sazba klesá.

Na službě BeeScale jsem si tedy prakticky vyzkoušel vytvořit a spustit server v cloudu. Po třídním testování musím konstatovat, že jsem se nesetkal s jediným problémem. Je zřejmé, že 256 MB paměti pro webový server není dostatečná hodnota, ale kromě ceny v tomto případě nic nebrání sestavit optimální řešení.

¹<https://beescale.com/front/kolik-to-stoji/>

4.5 Řešení *International Neuroinformatics Coordinating Facility (INCF)*

INCF je mezinárodní organizace, která podporuje a sjednocuje členské neuroinformatické skupiny z celého světa. Její snahou je mimo jiné vytvoření mezinárodní infrastruktury pro podporu výzkumu v oblasti neuroinformatiky. Sekretariát INCF sídlí ve Stockholmu ve Švédsku. Organizace sdružuje šestnáct zemí, mezi nimi i Českou Republiku. Výzkumná skupina KIV je členem českého národního uzlu. [17]

Organizace INCF byla postavena před problémem, jak navzájem sdílet informace z výzkumu členských skupin z oboru neuroinformatiků z různých částí světa. V rámci této podpory uvažovali o zakoupení výpočetního výkonu a datového úložiště od Amazone. Po zralé úvaze bylo v INCF rozhodnuto o vlastním datovém uzlu, který nese název DataSpace - celým názvem INCF DataSpace (IDS). Nejedná se o typický cloud storage, ale o vzájemné propojení datových úložišť členů organizace INCF. IDS tedy není služba dostupná pro širokou škálu běžných zákazníků, ale pouze pro výzkumníky z oblasti neuroinformatiky, kteří mají garantované bezplatné využití. IDS používá iRODS¹ software.

V současné době se INCF DataSpace zaměřuje na back-end úložné vrstvy. Byla zřízena datová iRODS základna pro celou federaci/unii. Jádro serveru je nasazeno v Amazone cloudu, je replikované přes čtyři zóny: západ USA, východ USA, Evropa a tichomořská část Asie. Tvoří tak jakousi páteř. Servery jsou dostupné na ids.incf.net, klienti budou přesměrováni na servery pro ně zeměpisně nejbližší.

Na páteř navazují servery, které poskytují úložiště nebo spravují další zóny. Tyto zóny pak komunikují v iRODS. Aktuální stav vývoje IDS lze sledovat online na Internetu², stejně tak jako dokumenty a dokumentace IDS³.

¹iRODS - integrated Rule-Oriented Data System

²<http://dev.incf.org/trac/infrastructure/wiki/CurrentResources>

³<http://dev.incf.org/trac/infrastructure/wiki/WikiStart>

5 Zhodnocení analyzovaných poskytovatelů

V předchozí kapitole byly analyzovány nabídky některých společností poskytující cloud computing. Společnost Amazon nabízí široké spektrum služeb cloud computingu, jak ostatně vyplývá z oddílu 4.1. Praktické vyzkoušení nabízených služeb ale nebylo možné realizovat z důvodu požadavku na údaje, které jsem nemohl společnosti Amazone poskytnout. Pokud by byly k dispozici finanční prostředky, bylo by řešení společnosti Amazone pro výzkumnou skupinu KIV (s ohledem na její požadavky) vhodné. Protože jsou ale veškeré služby i pro vědu, výzkum a školství plně zpoplatněny, nevyhovuje toto řešení požadavku na nízké (bezplatné) provozní náklady.

Oproti tomu společnost Google své cloudové služby poskytuje školství zdarma. Konkrétně se jedná o aktuálně volné výpočetní kapacity. Vzhledem ke skutečnosti, že neuroinformatická vědecká skupina KIV mimo jiné předpokládala pro svou EEG/ERP doménu stabilní cloudové řešení, a společnost Google pro cloud nabízí pouze své aktuálně volné kapacity, není cloud vhodným poskytovatelem pro danou výzkumnou skupinu. Další důvody vyplývající z analýzy řešení Google jsou uvedeny v oddílu 4.2.

Z analýzy nabídky společnosti Microsoft přichází v úvahu platforma Windows Azure. Z důvodu nedostatku referencí na toto řešení, ale také proto, že jsem jej nemohl prakticky otestovat kvůli dotazu na údaje o platební kartě, nelze řešení v současné chvíli akceptovat. Úložiště SkyDrive, které bylo možné prakticky vyzkoušet, nevyhovuje požadavkům výzkumné skupiny na kompatibilitu pro běhové prostředí Java.

Cloudová služba BeeScale je podle většiny požadavků výzkumné skupiny KIV optimální. Tuto službu bylo možné otestovat i bez zadání citlivých osobních dat. Virtuální server v cloudu je ovšem zpoplatněn, což nevyhovuje požadavku výzkumné skupiny KIV na nízké (bezplatné) provozní náklady.

Organizace INCF poskytuje softwarovou podporu výzkumu pro své členy zcela zdarma. Dle zjištěných informací přímo cloudovou službu výpočetního výkonu společnost nenabízí. Velmi zajímavé je ovšem řešení, které se nazývá INCF Dataspace. Služba by se dala popsat jako propojení datových úložišť členů INCF do společných uzlů. Považuji toto řešení za perspektivní již s ohledem na propojení s dalšími výzkumnými skupinami oboru neuroinformatiky a vzájemné sdílení výsledků výzkumu. Jak již bylo uvedeno, nejedná se o cloud ve „standardním“ smyslu. Nejdůležitějším požadavkem výzkumné skupiny na cloudové řešení byla

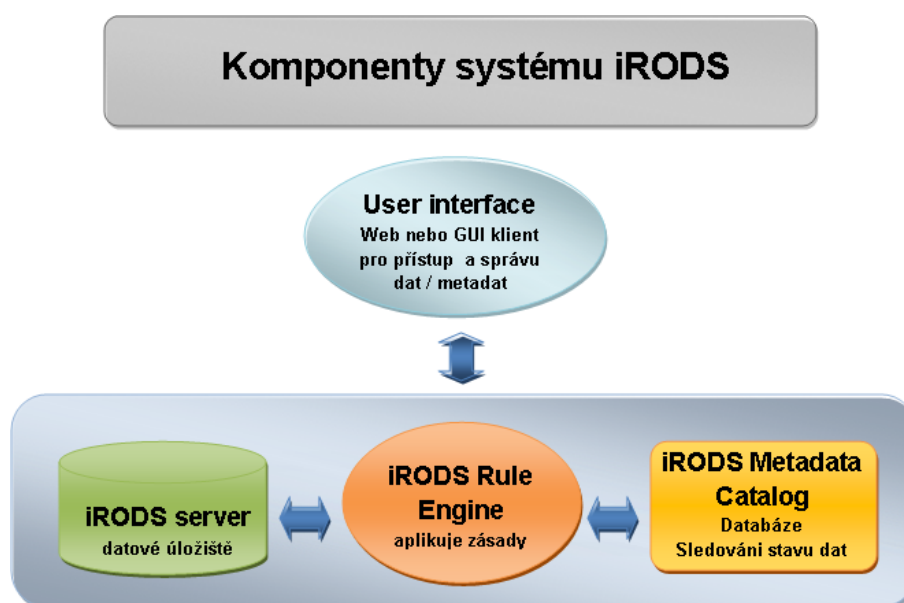
nízká finanční náročnost, proto po konzultaci s vedoucím práce bylo dohodnuto zprovoznění systému INCF Dataspace pro neuroinformatickou výzkumnou skupinu KIV.

6 INCF Dataspace a systém iRODS

Následující kapitola bude věnována systému iRODS.

6.1 Zakladní charakteristika iRODS

iRODS je systém, který pomáhá uživatelům organizovat, chránit i zachovávat velké množství souborů libovolné velikosti. Uživatelé mohou svá data nahrávat a sdílet s jinými uživateli. Všechna data jsou dostupná prostřednictvím jednotného systému. [18] Na obrázku č. 6 je znázorněno základní schéma složek, které tvoří systém iRODS.



Obrázek 6: Součásti systému iRODS

Rule Engine - aplikuje uživatelsky definované zásady a pravidla, prosazuje řídicí politiku.

iCAT Metadata Catalog - ukládá informace o stavu a popisná metadata do databáze, umožňuje vyhledávání, správu, řízení a sledování přístupu k datům.

Storage - úložný prostor (systém souborů) přístupný přes iRODS data server.

User interface - uživatelské rozhraní - přístup k serveru iRODS, viz. oddíl 6.2.

Zone - jedná se o uzel v systému, který se skládá ze serveru iRODS, databáze (iCAT) a příslušných datových zdrojů.

6.2 Klient systému iRODS

Systém iRODS je možné spravovat pomocí příkazů i-commands. [18] Typy příkazů a popisy jejich funkcí jsou k dispozici na internetu¹. Výhodou této administrace je jednoduchá instalace. Soubory i-commands pro operační systém Windows jsou též dostupné na internetu². V tomto případě postačí spustitelné soubory uložit na disk počítače a v příkazovém řádku zadávat příkazy i-commands s odpovídajícími parametry. Pro správné fungování příkazů je nezbytné spouštět soubory z jejich aktuálního adresáře, nebo mít cestu k nim uloženu v systémových proměnných. Přihlášení se provede zadáním příkazu `iinit`. Systém požádá o vložení přístupových údajů pro iRODS, tj. adresa, port, login a heslo. Po úspěšném ověření údajů je přihlášení na server provedeno a je možné systém dalšími příkazy administrovat nebo z něj získávat informace. Zadané parametry připojení (mimo hesla) si systém po úspěšném přihlášení uloží do souboru a při dalším připojení se tyto údaje nemusí znova zadávat.

Příklad vytvoření zóny:

```
iadmin mkzone cz.czu.eeg remote
```

Příklad přidání uživatele

```
iadmin mkuser testuser rodsadmin
```

Pro prohlížení dat v iRODS je k dispozici několik aplikací. Pro systém Windows je možné z webových stránek³ stáhnout iExplorer for Windows. Jedná se o rychlý navigační nástroj s bohatým grafickým uživatelským rozhraním. Navíc není nutné instalovat. Postačuje program (jeden exe soubor) uložit na disk počítače a spustit. Po spuštění se aplikace dotáže na přístupové údaje k serveru iRODS. Po jejich korektním zadání dojde k připojení na iRODS a je možné prohlížet a vyhledávat soubory. Na obrázku č. 7 je zobrazen náhled na aplikaci iExplorer.

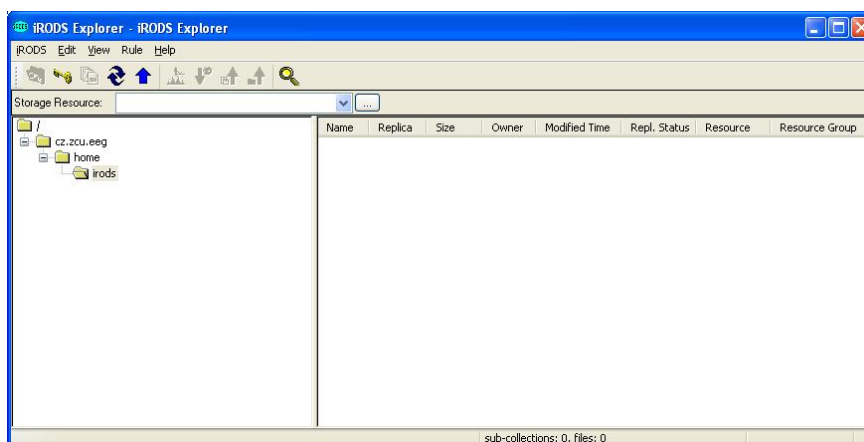
Jako další způsob prohlížení dat v iRODS je možné zmínit systém iDROP. [19] Jedná se o klienta postaveného na programovacím jazyku Java. K dispozici jsou tři verze této aplikace: iDROP Swing⁴ klient, iDROP webové rozhraní a iDrop-lite upload / download applet, který může být snadno vložen do vlastních webových rozhraní. iDROP webový klient umožňuje přístup z jakéhokoliv internetového prohlížeče (nutná podpora Java). Toto rozhraní je jedno z nejkompexnějších, po-

¹<https://www.irods.org/index.php/icommands>

²<https://www.irods.org/index.php/windows>

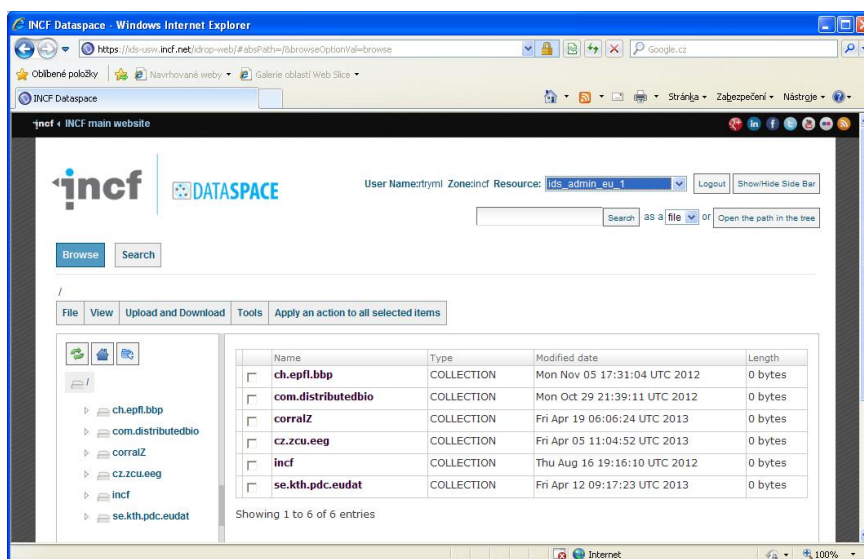
³<https://www.irods.org/index.php/windows>

⁴Swing - sada nástrojů grafického uživatelského rozhraní programovacího jazyka Java



Obrázek 7: Ukázka aplikace iExplorer for Windows

kud jde o editaci atributů objektů, metadat a ACL¹. Ukázka iDROP webového klienta je zobrazena na obrázku č 8.



Obrázek 8: Ukázka webového klienta iDrop

Pro úplnost je nutno uvést ještě dva způsoby přístupu do iRODS. Jsou jimi systém FUSE² a Davis (Webdav). FUSE - jedná se o Linuxovou knihovnu, která

¹ACL - Access control list (seznam pro řízení přístupu)

²FUSE - the Filesystem in User Space (Souborový systém v uživatelském prostoru)

umožňuje zapisovat do souborového systému. Protože tento systém obsahuje i jádro iRODS, příkazy, pomocí kterých se v Linuxu prochází souborový systém, fungují i v iRODS. Davis je webová aplikace založená na programovacím jazyku Java, která poskytuje rozhraní Webdav pro přístup do souborového systému iRODS podobně jako FUSE v Linuxu. Je možné také spravovat metadata a seznamy řízení přístupu ACL.

6.3 Připojení do systému INCF Dataspace

Postup připojení do IDS je uveden na internetu¹. [20] Instalační balíčky pro pro 64 bit. verzi Debian² Squeeze (6.0.4) jsou k dispozici v úložišti INCF. Jedná se o ids-tools, což jsou instalační a konfigurační skripty napsané v jazyku Python³. Před započítím instalace je nutné mít na operačním systému nainstalovány potřebné moduly:

- python-pip
- python-crypto

Moduly lze do systému Debian přidat pomocí správce balíčků, např. Aptitude, nebo příkazem `apt-get -f install`. Po úspěšném nainstalování je možno přistoupit ke spuštění příkazu:

```
sudo pip install http://ids-us-east-1.s3.amazonaws.com/ids-tools-latest.tar.gz
```

pip je nástroj pro instalaci a správu balíčků Python. Umožňuje snadné přidání balíčků skriptů příkazem `pip install [název balíčku]` i jeho následné jednoduché odebrání příkazem `pip uninstall [název balíčku]`. Instalaci balíčků je nutno provádět pod uživatelem s administrátorskými právy. Případně uživatel musí být členem skupiny sudo⁴. Pokud instalace ids-tools proběhne úspěšně, je možné spouštět jednotlivé instalační skripty.

Následující popis je věnován postupu přidání zóny do federace INCF. Předpokladem je funkční iRODS server včetně iCAT databáze. V INCF je dohodnuto,

¹<https://github.com/INCF/ids-tools/wiki/IDS-Zone-Setup>

²Debian - distribuce operačního systému Linux, na kterém nainstalován iRODS výzkumné skupiny

³Python - dynamický, objektově orientovaný skriptovací programovací jazyk. Standardně nainstalován v systému v rámci distribuce Debian

⁴sudo - substitute user do - slouží k vykonání operace pod jiným uživatelem, v tomto případě například uživatelem root

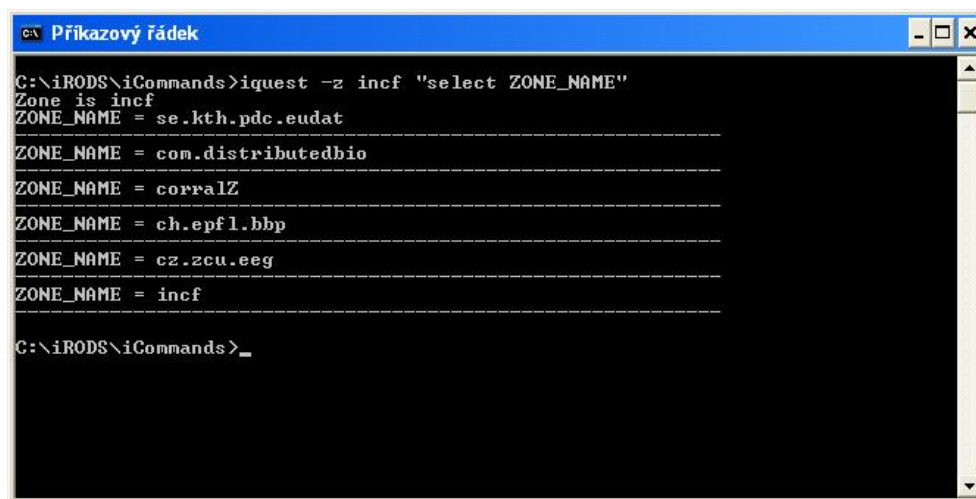
že v rámci federace je název zóny obrácený název internetové domény příslušné výzkumné skupiny. Pokud má laboratoř název internetové domény např. "lab.univerzity.edu", měla by mít příslušná zóna laboratoře název "edu.univerzity.lab". Dalším předpokladem je tedy na iRODS vytvořena zóna s názvem cz.zcu.eeg. Je znám login a heslo administrátora iRODS serveru. Příkazem iCommands `iinit` se připojíme k administraci iRODS serveru. Spustíme příkaz:

```
sudo -i iadmin mkzone incf remote ids.incf.net:1247
```

Tímto příkazem dojde k připojení zóny `incf` k iRODS serveru. V dalším kroku je nutné poslat emailovou zprávu na adresu `ids-admin@incf.org` s údaji iRODS (iCAT) serveru. To jest FQDN¹ a port, na kterém běží iRODS server (standardně se jedná o port číslo 1247). Stejným způsobem, jako byla přidána zóna `incf` do našeho iRODS serveru, přidá administrátor INCF zónu `cz.zcu.eeg` do jejich systému a pošle emailovou zprávu, že tak bylo učiněno a vše proběhlo korektně. Že tomu tak skutečně je, je možno ověřit tímto způsobem: po přihlášení k systému iRODS pomocí `iinit` se zadá příkaz iCommands:

```
iquest -z incf "select ZONE_NAME"
```

Zobrazí se výpis dostupných zón iRODS serveru, viz obrázek č. 9.



```
ca Příkazový řádek
C:\iRODS\iCommands>iquest -z incf "select ZONE_NAME"
Zone is incf
ZONE_NAME = se.kth.pdc.eudat
-----
ZONE_NAME = com.distributedbio
-----
ZONE_NAME = corralZ
-----
ZONE_NAME = ch.epfl.bbp
-----
ZONE_NAME = cz.zcu.eeg
-----
ZONE_NAME = incf
-----
C:\iRODS\iCommands>_
```

Obrázek 9: Výpis seznamu zón na iRODS serveru

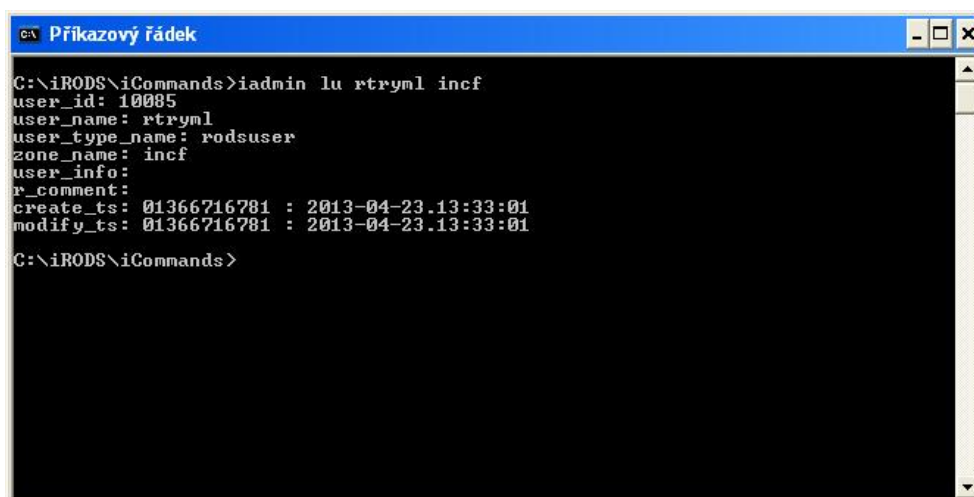
¹FQDN - Fully Qualified Domain Name - je označení pro plně specifikované doménové jméno počítače.

Pro sdílení dat v rámci federace INCF je zapotřebí získat seznam uživatelských přístupů. Čili uživatelské účty zóny incf je nutné synchronizovat (vytvořit je) i v zóně incf na iRODS serveru neuroinformatické výzkumné skupiny. K tomuto účelu slouží script `ids-sync-users` balíčku `ids-tools`. Pro spuštění stahování uživatelských účtů pro zónu incf se zadá příkaz:

```
sudo -i ids-sync-users
```

Úspěšné vytvoření uživatelských přístupů zóny incf na iRODS serveru je možné ověřit například příkazem `iCommands`:

```
iadmin lu [login] [zóna]
```



```
C:\iRODS\iCommands>iadmin lu rtryml incf
user_id: 10085
user_name: rtryml
user_type_name: rodsuser
zone_name: incf
user_info:
r_comment:
create_ts: 01366716781 : 2013-04-23.13:33:01
modify_ts: 01366716781 : 2013-04-23.13:33:01
C:\iRODS\iCommands>
```

Obrázek 10: Výpis informace o uživateli v zóně incf

Nyní je možné používat skupiny a uživatele zóny incf pro nastavení přístupu ve vlastní zóně `cz.zcu.eeg`. Pokud jsou tedy tato oprávnění nastavena, může uživatel zóny incf, který má oprávnění čtení, v rámci `cz.zcu.eeg` procházet soubory zóny `cz.zcu.eeg`. Aby se udržoval seznam uživatelů zóny incf aktuální, je vhodné spouštět příkaz `ids-sync-user` pravidelně například automaticky z cronu¹.

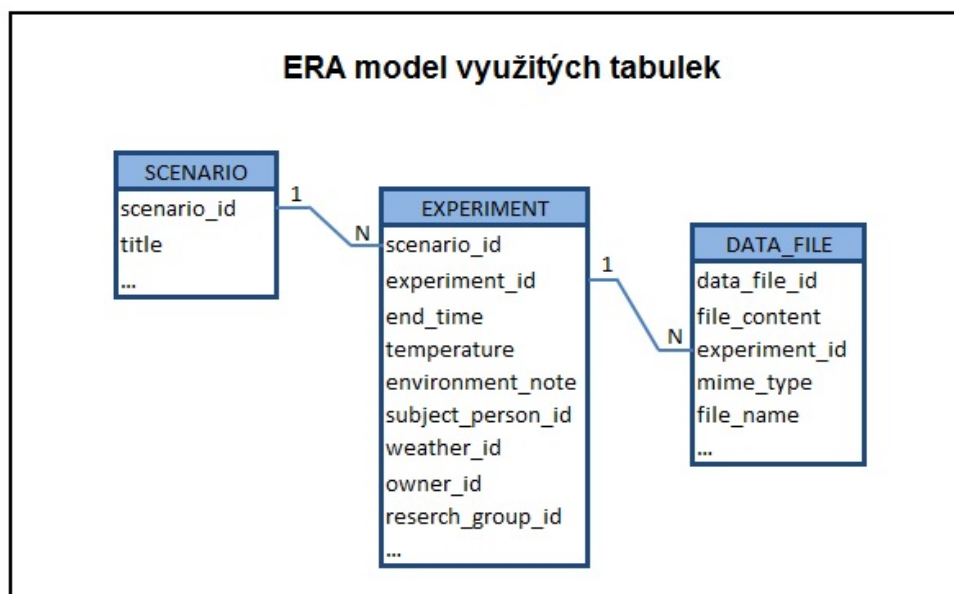
¹cron - softwarová služba, které v operačních systémech spouští v daný čas určitý příkaz, script atd.

7 Získání dat z databáze EEG pro INCF Dataspace

Tato kapitola pojednává o extrakci dat z databáze neuroinformatické skupiny pro systém Dataspace.

7.1 Úvodní informace

Server iRODS (system Dataspace) využívá pro data souborový systém. Je proto nutné vytvořit program pro získání potřebných dat z databáze Oracle. Tato data budou následně uložena do souborů. Pro systém Dataspace bude podle zadání vytvořena následující struktura: z tabulky scénářů (SCENARIO) je nutné získat všechny názvy daných scénářů. Tyto názvy budou tvořit složky. V každé složce se scénářem budou uloženy složky s názvem dle čísla experimentu (z tabulky EXPERIMENTS) příslušné k danému scénáři. Pro názornost je na obrázku č. 11 uvedeno schéma tabulek.



Obrázek 11: Struktura využitých tabulek z databáze EEG

V každé složce experimentu bude uložen archiv ZIP s odpovídajícími soubory z tabulky DATA_FILE příslušných danému experimentu. Ve složce experimentu bude dále uložen soubor (např. csv) s informacemi o datech (tzv. metadata). Jedná se o další údaje o jednotlivých experimentech, například datum a čas začátku i konce experimentu, teplota, počasí atd. Tento vytvořený souborový systém bude

uložen v adresáři s názvem *experiments* na sereru iRODS pro sdílení dat v rámci systému Dataspace.

7.2 Řešení pomocí jazyka PL/SQL

Data z výzkumu neuroinformatické skupiny jsou uložena v databázi Oracle 11g. Pro práci s databází Oracle je vhodný jazyk PL/SQL¹. V PL/SQL jsou použity konstrukce známé především z vyšších programovacích jazyků: proměnné a konstanty, cursory, cykly a rozhodovací podmínky, subprogramy atd. Implementace v jazyce PL/SQL byla zvolena jako první varianta řešení. Důvodem byla zmíněná komplexnost jazyka PL/SQL nad databází Oracle a dále pak proto, že skripty PL/SQL lze spouštět interaktivně z webových stránek (např. portálu EEG). Od tohoto řešení bylo ale prozatím upuštěno, protože nebyl nalezen způsob získání BLOB z databáze. Řešením by bylo ukládat data z BLOB pomocí balíčku UTL_FILE. UTL_FILE lze v jazyce PL/SQL použít pro zápis do lokálního souborového systému, kde je umístěna databáze. [21] Protože přístup do souborového systému databáze EEG není možný, bylo od této možnosti řešení upuštěno.

Dalším možným řešením bylo posílat data z BLOB protokolem TCP přes síťové spojení na server iRODS. Na serveru iRODS by běžel program, který by posílaná data zpracovával a ukládal do lokálního souborového systému na serveru. K realizaci toho řešení lze použít balíček UTL_TCP. [22] Protože toto řešení bylo implementačně složité (obsahuje dvě části, jednu na databázi a druhou na serveru iRODS) a navíc by se musely obě části udržovat ve vzájemné synchronizaci, bylo zvoleno řešení v jazyce Java.

7.3 Řešení pomocí jazyka Java

Programovací jazyk Java obsahuje podporu pro připojení k databázi. Pro každou databázi, se kterou chtějí programy v jazyce Java komunikovat, musí být ale vytvořen ovladač tzv. JDBC² driver. Pro databázi Oracle je JDBC ovladač k dispozici ke stažení z webových stránek³ jako `ojdbc6.jar` soubor pro JDK 1.6. Pro získání dat z databáze v jazyce Java jsou důležité třídy balíku `java.sql`. Třída *Connection* slouží pro vytvoření spojení k databázi. Třída *PreparedStatement* pro zaslání SQL dotazu/příkazu na databázi a *ResultSet* pro uložení výsledků dotazu.

¹PL/SQL - Procedural Language/Structured Query Language - je procedurální nadstavba jazyka SQL pro databáze Oracle

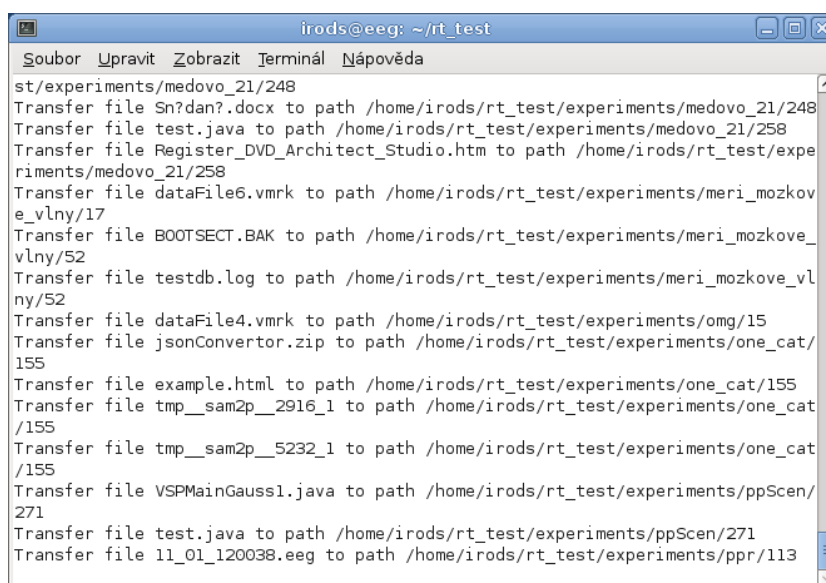
²JDBC - Java Database Connectivity - aplikační rozhraní, které definuje jednotný přístup k relačním databázím

³<http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html>

Algoritmus programu je následující: po spuštění dojde z konfiguračního souboru k přečtení přístupových údajů k databázi a vytvoří se spojení. Pomocí SQL dotazu z tabulky SCENARIO se vyberou všechny názvy scénářů a v cyklu se prochází každý scénář zvlášť. Ke každému scénáři se opět pomocí SQL dotazu získají všechny experimenty z tabulky EXPERIMENTS a v cyklu se každému experimentu vyberou pomocí SQL dotazu všechny odpovídající soubory z tabulky DATA_FILE. Pro čtení souborů se vytvoří čtecí proud binárních dat. Pro zápis dat do souboru program vytvoří výstupní proud třídy ZIPOutputStream. Tímto způsobem dochází k archivaci všech souborů daného experimentu do jednoho ZIP archivu. V posledním kroku dojde k získání metadat o příslušném experimentu. Metadata se uloží do souboru csv ve složce daného experimentu. Charakter metadat lze ovlivnit SQL dotazem. Ve standardním nastavení programu se do csv souboru ukládají všechny položky příslušného experimentu viz obrázek č. 11, tabulka EXPERIMENT. [23]

7.4 Zhodnocení řešení

Program byl testován na operačním systému Windows 7 a Linux Debian. Průběh činnosti programu je znázorněn na obrázku č. 12.



```
irods@eeg: ~/rt_test
Soubor Upravit Zobrazit Terminál Nápověda
st/experiments/medovo_21/248
Transfer file Sn?dan?.docx to path /home/irods/rt_test/experiments/medovo_21/248
Transfer file test.java to path /home/irods/rt_test/experiments/medovo_21/258
Transfer file Register_DVD_Architect_Studio.htm to path /home/irods/rt_test/experiments/medovo_21/258
Transfer file dataFile6.vmrk to path /home/irods/rt_test/experiments/meri_mozkove_vlny/17
Transfer file BOOTSECT.BAK to path /home/irods/rt_test/experiments/meri_mozkove_vlny/52
Transfer file testdb.log to path /home/irods/rt_test/experiments/meri_mozkove_vlny/52
Transfer file dataFile4.vmrk to path /home/irods/rt_test/experiments/omg/15
Transfer file jsonConvertor.zip to path /home/irods/rt_test/experiments/one_cat/155
Transfer file example.html to path /home/irods/rt_test/experiments/one_cat/155
Transfer file tmp_sam2p_2916_1 to path /home/irods/rt_test/experiments/one_cat/155
Transfer file tmp_sam2p_5232_1 to path /home/irods/rt_test/experiments/one_cat/155
Transfer file VSPMainGauss1.java to path /home/irods/rt_test/experiments/ppScen/271
Transfer file test.java to path /home/irods/rt_test/experiments/ppScen/271
Transfer file 11_01_120038.eeg to path /home/irods/rt_test/experiments/ppr/113
```

Obrázek 12: Ukázka práce programu na výběr dat z databáze EEG

Vyzkoušel jsem aplikaci i na serveru v cloudu ve službě BeeScale. Rychlost stahování dat z databáze EEG byla i případě cloudového řešení totožná s rychlostí

stahování na serveru iRODS. Z důvodu, že se jedná o neveřejná data výzkumné skupiny KIV byl tento test proveden pouze na několika souborech. Stažené soubory byly následně z cloudového serveru odstraněny. V případě, že by tato data byla skutečně uložena na serveru v cloudu, musela by se nezbytně řešit problematika šifrování.

8 Závěr

Zadání bakalářské práce bylo nálezt optimální cloudové řešení pro EEG/ERP doménu. Byly analyzovány požadavky neuroinformatické skupiny KIV a na jejich základě byl proveden průzkum dostupných řešení, z nichž bylo vybráno takové, které nejvíce vyhovovalo požadavkům výše uvedené skupiny.

V současné době je pro výzkumnou skupinu KIV vhodný systém Dataspace, který vyvíjí a podporuje nezisková organizace INCF. Systém je postaven na technologii iRODS. Jedná se o technologii sdílení dat z výzkumu mezi jednotlivými členskými skupinami. V tomto smyslu je tato technologie přínosná, protože spojuje členské laboratoře po celém světě do jednoho celku. Ačkoliv na první pohled nejde o pronájem výpočetních kapacit od poskytovatele cloudu, je nutné se na tuto záležitost dívat v širším slova smyslu. Výzkumná skupina KIV produkuje a ukládá informace ze svých vlastních výzkumů, k dispozici má oproti tomu výsledky výzkumu členských laboratoří z celého světa.

V současné době je systém Dataspace funkční a data neuroinformatické skupiny určena pro sdílení jsou v zóně incf přístupná. Také soubory, které jsou sdíleny dalšími laboratoroři, jsou díky systému Dataspace pro výzkumnou skupinu KIV k dispozici. Program pro výběr souborů a metadat z databáze EEG je možné pravidelně spouštět tak, aby data výzkumné skupiny byla v systému Dataspace vždy aktuální. Jako vylepšení této skutečnosti je možná úprava programu tak, aby jej bylo možné spouštět přímo, například z webových stránek portálu EEG/ERP.

Zkušenost s hledáním cloudového řešení dle zadaných požadavků hodnotím jako přínosnou. Jelikož jsem se s tímto odvětvím informatiky doposud blíže nezabýval, mnoho informací pro mě bylo úplnou novinkou. Bylo potřeba prostudovat velké množství materiálů a zároveň ověřit řadu faktů. Pozitivně též hodnotím, že jsem se mohl seznámit s technologiemi, které využívá neuroinformatická skupina KIV, aktivně jsem komunikoval s organizací INCF a řešil odborné problémy s napojením iRODS serveru do systému Dataspace. Naučil jsem se pracovat s databázovými systémy Oracle, které využívá zmíněná výzkumná skupina. Všechny tyto nově nabyté informace jsou pro mne důležité nejen v zaměstnání, ale také v dalším případném studiu.

Přehled zkratk

ACL - access control list - seznam oprávnění přístupu k objektu (např. souboru)

API - Aplikační programové rozhraní

AWS - Amazon Web Services

EEG - Elektroencephalography - vyšetření registrující elektrickou aktivitu mozku

ERP - Even-Related Potential - Evokovaný potenciál - mozková odezva na stimulaci

FQDN - Fully Qualified Domain Name - doménový název počítače nebo serveru

iRODS - integrated Rule-Oriented Data System

IT - informační technologie

JDBC - Java Database Connectivity - aplikační rozhraní aplikací Java pro přístup k databázi

KIV - Katedra informatiky a výpočetní techniky

PL/SQL - Procedural Language/Structured Query Language - procedurální nadstavba jazyka SQL od firmy Oracle

SQL - Structured Query Language - standardizovaný dotazovací jazyk používaný pro práci s daty v relačních databázích

SSD - Solid-State Drive - datové médium, neobsahuje mechanické části, vyšší přenosové rychlosti

ZČU - Západočeská univerzita v Plzni

Literatura

Tištěné články:

HEROUT, Pavel. *Učebnice jazyka Java*. 3., rozš. vyd. České Budějovice: Kopp, 2007, 381 s. ISBN 978-80-7232-323-4.

URMAN, Scott, Ron HARDMAN a Michael MCLAUGHLIN. *Oracle: programování v PL/SQL*. Vyd. 1. Brno: Computer Press, 2007, 720 s. ISBN 978-80-251-1870-2.

Elektronické články:

[1] JEŽEK, Petr a Roman MOUČEK. *Database of EEG/ERP experiments*. 2009.

[2] *Cloud computing* [online]. 2013 [cit. 24.4.2013].
Dostupné z: http://cs.wikipedia.org/wiki/Cloud_computing

[3] *Amazon Web Services* [online]. 2013 [cit. 26.4.2013].
Dostupné z: <http://aws.amazon.com/>

[4] *Amazon Elastic Compute Cloud (Amazon EC2): Cloud Computing Servers*. [online]. [cit. 26.4.2013].
Dostupné z: <http://aws.amazon.com/ec2/>

[5] *Amazon Simple Storage Service (Amazon S3)* [online]. 2013 [cit. 26.4.2013].
Dostupné z: <http://aws.amazon.com/s3/>

[6] *AWS Free Usage Tier* [online]. 2013 [cit. 26.4.2013].
Dostupné z: <http://aws.amazon.com/free>

[7] *Google Apps pro vzdělávání* [online]. 2012 [cit. 26.4.2013].
Dostupné z: <http://www.google.com/apps/intl/cs/edu/>

[8] *Google App Engine* [online]. 2013 [cit. 27.4.2013].
Dostupné z: <http://appengine.google.com/>

[9] *Google Compute Engine* [online]. 2012 [cit. 27.4.2013].
Dostupné z: <https://developers.google.com/compute/>

[10] *Google Cloud Storage* [online]. 2013 [cit. 27.4.2013].
Dostupné z: <https://cloud.google.com/products/cloud-storage>

[11] *Google BigQuery* [online]. 2013 [cit. 27.4.2013].
Dostupné z: <https://cloud.google.com/products/big-query>

- [12] *Google Cloud SQL* [online]. 2013 [cit. 27.4.2013].
Dostupné z: <https://cloud.google.com/products/cloud-sql>
- [13] *Google končí s podporou protokolu Exchange ActiveSync, RIP Google Sync* [online]. 2012 [cit. 27.5.2013].
Dostupné z: <http://www.lupa.cz/clanky>
- [14] *Windows Azure: Veřejný cloud Microsoft*. [online]. [cit. 26.4.2013].
Dostupné z: <http://www.microsoft.com/cs-cz/server-cloud/windows-azure.aspx>
- [15] *Microsoft SkyDrive: Microsoft Windows*. [online]. [cit. 26.4.2013].
Dostupné z: <http://windows.microsoft.com/cs-cz/skydrive>
- [16] *BeeScale* [online]. 2013 [cit. 27.5.2013].
Dostupné z: <https://beescale.com/>
- [17] *INCF* [online]. 2012 [cit. 27.5.2013].
Dostupné z: <http://www.incf.org/>
- [18] *IRODS* [online]. 2013 [cit. 27.5.2013].
Dostupné z: <http://www.irods.org>
- [19] System Architecture and Design, Data Management Policy Specification. *Dataspace* [online]. 2012, 1.3 [cit. 27.4.2013]. Dostupné z: <http://www.incf.org/resources/data-space/documentation>
- [20] *IDS Zone Setup* [online]. 2013 [cit. 27.4.2012].
Dostupné z: <https://github.com/INCF/ids-tools/wiki/IDS-Zone-Setup>
- [21] *How can I extract files from an Oracle BLOB field?*[online]. 2011 [cit. 26.4.2013].
Dostupné z: <http://stackoverflow.com/questions/6332032/how-can-i-extract-files-from-an-oracle-blob-field>
- [22] *UTL_TCP* [online]. 2007 [cit. 27.4.2013].
Dostupné z: http://docs.oracle.com/cd/B19306_01/appdev.102/b14258/u_tcp.htm
- [23] *Read BLOBs data from database* [online]. 2012 [cit. 26.4.2013].
Dostupné z: <http://www.java2s.com/Code/Java/Database-SQL-JDBC/ReadBLOBsdatafromdatabase.htm>

Příloha č. 1

Výběr dat z databáze EEG pro systém Dataspace

Uživatelská dokumentace

1) Účel programu:

Program slouží pro výběr dat z databáze. Konkrétně se jedná o získání souborů a metadat z databáze Oracle neuroinformatické výzkumné skupiny KIV pro systém Dataspace.

2) Popis programu:

Program byl vyvinut v programovacím jazyku Java. Jeho činnost je jednorázová a automatická. Není tedy interaktivně závislá na ovládní uživatelem. Po spuštění se program napojí na databázi EEG. Vybere postupně scénáře z databáze a také ke každému scénáři všechny příslušné experimenty. Ke každému experimentu program nalezne všechny odpovídající soubory. V dalším kroku program vybere z databáze informace o každém experimentu a uloží je do zvláštního souboru. O své aktuální činnosti, chybách a úspěšném ukončení program informuje výpisy na obrazovce.

3) Instalace

Jak již bylo uvedeno, program je vyvinut v programovacím jazyku Java. Je k dispozici ve zdrojovém kódu pro kompilaci na libovolném operačním systému, na kterém běží Java virtuální stroj. Pro operační systém Windows je zdrojový kód již zkompilován pro verzi Java 7 a je vytvořen spustitelný archiv JAR se všemi potřebnými knihovnamí. Identicky je program sestaven pro operační systém Debian s verzí Java 6. Pro snazší sestavení na jiném operačním systému je program k dispozici jako projekt Maven. Při kompilaci zdrojového kódu musí být překladači k dispozici informace o uložení souboru ojdbc6.jar (ovladač pro databázi Oracle). Ještě před spuštěním programu je potřeba ověřit, zda je ve stejném adresáři s programem uložen konfigurační soubor s přístupovými údaji k databázi. Pokud tomu tak nebude, program skončí s chybou. První řádek konfiguračního souboru udává položku host. Druhý údaj představuje port. Třetí řádek má hodnotu Service ID (SID). Čtvrtý a pátý řádek konfiguračního souboru udává login a heslo.

4) Spuštění programu

Program se spouští z konzole nebo příkazového řádku příkazem:

```
java -jar EEGTransfer.jar
```

V tomto případě v aktuálním adresáři, odkud je program spuštěn, se vytvoří složka *experiments*, ve které se budou vytvářet složky se scénáři, experimenty a soubory. Pokud by byl požadavek na změnu výstupního adresáře, program očekává jeden parametr s úplnou cestou pro uložení dat. Syntaxe takového příkazu by pak byla například:

```
java -jar EEGTransfer.jar ./home/irods/
```

V uvedeném adresáři se založí složka *experiments*, ve které se pak budou vytvářet adresáře a stahovat soubory identickým způsobem, jako je uvedeno výše. Program stahované soubory přímo komprimuje do ZIP archivů. Doba běhu programu závisí na velikosti přesouvaných dat a rychlosti síťového spojení. Z tohoto důvodu běh programu trvá řádově několik minut (16 minut transport dat z databáze EEG na server iRODS).

5) Chyby programu a jejich řešení:

Program je dostatečně otestován, přesto nelze vyloučit při spuštění a běhu programu k výskytu chyb, které ale budou zachyceny systémem vyjímek. Jedná se zejména o chyby čtení a zápisu do souborů (tzv. *IOException*) a chyby v souvislosti s operacemi na databázi (tzv. *SQLException*). Následuje seznam některých chybových hlášení a možný způsob jejich odstranění:

Error reading data access to the database. - Nedostupný soubor s konfiguračními údaji k databázi, nebo z něj nelze číst. Nutno prověřit dostupnost konfiguračního souboru.

Error initializing database driver. - Nedostupný soubor *ojdbc6.jar* (ovladač pro databázi Oracle) v cestě, která je uvedena v seznamu knihoven, nebo závislostí programu.

Error connecting to database. - Chyba spojení s databází. V konfiguračním souboru jsou uvedeny nesprávné údaje pro přístup k databázi. Chyba sítě.

Error the file transfer from database. - Chyba kopírování souborů z databáze. K chybě dochází buď v případě čtení binárních dat z datového objektu BLOB v databázi, nebo při zápisu do souborového systému na serveru iRODS.

Transfer file - SQL Error. - K této chybě dochází, pokud selže SQL dotaz na databázi. Například v databázových tabulkách došlo k modifikaci, kterou program nepředpokládal. Je třeba upravit SQL dotaz dle nového uspořádání tabulek.

Create metadata - SQL Error. - Identická chyba jako výše, ale vznikla při vytváření metadat.

Create metadata - File create error. - Chyba při zápisu do souborů s metadaty na serveru iRODS. Není povolen zápis, aktuální soubor je pro zápis uzamčen nebo soubor má otevřen jiný proces.