



Publikováno na *Ikaros* (<https://ikaros.cz>)

[Domů](#) > Virtualizovaná databáza vedeckej korešpondencie v rokoch 1500-1800 s využitím softvéru DSpace

Virtualizovaná databáza vedeckej korešpondencie v rokoch 1500-1800 s využitím softvéru DSpace

0 comments

Podnázov: Virtualizovaný DSpace použitý inak

[Anglicky](#)

English title:

Virtualized database of scientific correspondence between the years 1500-1800 using DSpace software

English subtitle: Virtualized DSpace used in a different way

English abstract:

Virtualization today is considered as a common technology that is increasingly being used outside an enterprise segment, for example in educational, research, and memory institutions. Thanks to several free available open source solutions, this technology is available for the institutions in which low or no license fees for software use are considered preferable. We are also describing the situation in which a virtualized system with DSpace is used in a different, alternative way - to store and access metadata records of various entities. Our case study is based on the practical implementation of a virtualized online database of european scientific correspondence dated between the years 1500-1800.

Autoři: [Formanek, Matúš](#) ^[1]

Vydání: [2019, ročník 23, číslo 2](#) ^[2]

Rubrika: [Pořádání informací](#) ^[3]

Úvod k virtualizácii serverov

Pod pojmom virtualizácia chápú **Singh** a **Chu** (2016) v širšom význame oddelenie zdrojov alebo požiadaviek od fyzického poskytovania týchto služieb. V užšom zmysle slov ide o koncept, v ktorom sa prístup ku konkrétnemu hardvérovému zariadeniu koordinuje spôsobom, pri ktorom sa viaceré operačné systémy môžu o tieto zdroje navzájom deliť. Johnston s kolektívom (2013) vnímajú virtualizáciu ako proces vytvárania a riadenia virtuálnych verzií týchto zdrojov alebo entít. Typickú entitu predstavuje operačný systém alebo hardvérový komponent, napr. sieťová karta.

V prípade klasického, bežného riešenia, je nad fyzickým hardvérom nainštalovaný zvyčajne jeden operačný systém so všetkými potrebnými službami. V prípade virtualizovaného riešenia sa jedná o modernizáciu klasického modelu prostredníctvom virtualizačnej vrstvy vlozenej medzi základný fyzický hardvér a ostatné softvérové komponenty. Tento moderný prístup pri implementácii systémov považujú viacerí autori (napr. Turner, Barron a McCrocklin 2016) medzi najviac inovatívne evolúcie v IT technológiách. Trend virtualizácie ponúka mnohé nové, široké spôsoby využitia i vo vedeckovýskumných inštitúciách – či už v testovacom výskumnom alebo produkčnom prostredí.

Virtuálny server vzniká nad fyzickým hardvérom, s ktorým veľmi tesne, t.j. na hardvérovej úrovni komunikuje tzv. *virtualizačný hypervízor* (Turner, Barron, McCrocklin 2016), ktorý Singh a Chu (2016) definujú hypervízor ako monitorovací nástroj virtuálnych strojov. Vrstva hypervízora umožňuje viacerým operačným systémom zdieľať jednotný hosťateľský hardvér. Úloha

hypervízora, resp. celej virtualizačnej vrstvy spočíva ďalej v emulovaní, kontrolovaní a synchronizovaní alokovaných hardvérových zdrojov voči hosťovaným systémom podľa aktuálnych požiadaviek. Prvé pokusy s virtualizáciou sa objavili už na prelome 60. a 70. rokov 20. storočia v spoločnosti IBM. Masovo sa tento druh softvéru začal predávať okolo roku 2000 (Singh a Chu 2016). V súčasnosti dosahuje vysoký stupeň vývoja a prináša mnoho inovácií, ako i nesporných výhod do IT prostredia.

Výhody virtualizácie

Ako sme už spomínali, pri klasickom riešení servera je na jednom stroji nainštalovaný zvyčajne jeden operačný systém, ktorý priamo využíva prostriedky hardvéru. Nad operačným systémom sú následne spustené všetky nevyhnutné aplikácie, databázy, webové servery a používateľské rozhrania tvoriace súčasti virtuálneho priestoru (Turner, Barron a McCrocklin 2016). Virtualizácia naopak prináša úplnú nezávislosť operačného systému a použitých aplikácií od hardvéru. Viaceré súbežne spustené operačné systémy spoločne s aplikáciami môžeme zapuzdriť do oddelených virtuálnych strojov (Singh a Chu 2016). V každej inštancii virtuálneho stroja tak vzniká potenciálne samostatný virtuálny priestor schopný koexistovať s ostatnými sesterskými priestormi v istej súčinnosti alebo celkom nezávisle.

Turner, Barron a McCrocklin (2016) ďalej medzi podstatné výhody virtualizácie radia aj úspory nákladov na hardvérové komponenty, zvýšenie stability riešenia, ako aj význačné aspekty škálovateľnosti celej infraštruktúry. Vďaka virtualizovanému prostrediu môžeme vytvárať (či rušiť) jednotlivé virtuálne entity podľa aktuálnych potrieb. Rovnako tak je možné dynamicky alokovať, a tým pádom i škálovať väčšinu zdrojov priradenú týmto entitám. V dôsledku toho dosiahneme použitím virtualizácie vo výsledku vyšší stupeň efektívnosti, t. j. zníženie potrebných zdrojov (Turner; Barron; McCrocklin 2016). Konsolidácia nastane vďaka tomu, že rozličné aplikácie spustíme na viacerých virtuálnych strojoch, ktoré síce fungujú oddelene, no na jednom, výkonnejšom, fyzicky zdieľanom stroji (Spink; Wagstaff; Franke 2016).

Singh a Chu (2016) uvádzajú, že pri nasadení virtualizácie serverov je možné dosiahnuť mieru úspory zdrojov na úrovni od 40% do 80%. Logickým dôsledkom už spomenutého sa stáva zjednodušenie správy týchto strojov – a to ako hosťovateľského fyzického servera, tak i virtuálnych strojov. V prípade potreby je možné migrovať virtuálne stroje na iný hosťovateľský server využívajúci podobnú virtualizačnú technológiu. A to aj v prípade, že sa fyzický hardvér druhého servera odlišuje, resp. je novší, modernejší. Jednoduchým spôsobom tak môžeme modernizovať fyzickú infraštruktúru, navyše bez dlhších výpadkov spustených služieb (Singh a Chu 2016).

Na druhej strane, značná nevýhoda virtualizácie vyplýva z vyššej spotreby hardvérových prostriedkov – procesora, pamäti a vstupno-výstupných operácií – v porovnaní s fyzickými servermi pri porovnateľnej záťaži (Singh a Chu 2016). Virtualizačný hypervízor totiž sám osebe spotrebuje časť výkonu na vlastnú réžiu a správu aktuálne spustených virtuálnych strojov. V inej našej štúdii (Formanek 2018) sme však načrtli, že strata výkonu sa pohybuje v priemere len v jednotkách percent.

Dostupné virtualizačné platformy

Mnoho svetových univerzít v posledných rokoch zavádza prostriedky virtualizácie aj do svojho akademického prostredia, čo možno badať na vzniku viacerých nedávno publikovaných diel venovaných hypervízorom od spoločnosti VMWare (Turner, Barron a McCrocklin 2016). Práve táto spoločnosť patrí dnes k vedúcim lídrom v oblasti virtualizácie so svojim hypervízorom ESXi určeným pre profesionálny enterprise segment. Ide však o proprietárny produkt, ktorého možnosti správy sú bez platenej licencie značne obmedzené. Manažment prevádzky viac ako jedného servera súčasne sa stáva prakticky nemožným. Podobne tak riešenie od Microsoftu v podobe Hyper-V Servera patrí medzi produkty s podobným typom licencovania. Kedysi bolo Hyper-V súčasťou produktu Microsoft Server. Dnes sa dodáva zdarma ako samostatné riešenie založené na jadre Microsoft Server.

K plnohodnotnému použitiu a pokročilej správe virtuálnych strojov je však nutné jednotlivé inštancie riadiť prostredníctvom dedikovanej (a platenej) inštancie Windows Servera. Aj to sú dôvody, kvôli ktorým nepokladáme tieto riešenia za tak zaujímavé vzhľadom na vyššie nadobúdacie náklady, ktoré môžu v istých situáciách predstavovať problém. Centrum záujmu predstavujú v našom prípade open source alternatívy. V našej konkrétnej štúdií sme sa rozhodli pre experimentálnu implementáciu voľne dostupného riešenia Citrix XenServer. Hlavným dôvodom bolo stabilné zázemie, široká používateľská základňa a nulové obstarávacie náklady pri použití verzie XenServera bez špeciálnej platenej licencie. Táto *free* verzia síce nedisponuje pokročilými funkciami, no i tak postačuje na zabezpečenie funkcionalít kladených na systém ako celok.

Implementácia virtualizovanej databázy INDED

Predložená kapitola stručne popisuje riešenie prípadovej štúdie, ktorej cieľom bola implementácia online databázy vedeckej korešpondencie vybraných európskych vzdelancov v rokoch 1500-1800. Databáza predstavuje jeden z výstupov projektu APVV-15-0554 s názvom *Intelektuálne dedičstvo a vedecká komunikácia 1500-1800 so slovenskými vzťahmi ako súčasť európskej histórie a identity*, skrátene INDED.

Databázou INDED v tomto kontexte myslíme určitý *ad hoc* interaktívny elektronický systém dostupný online, ktorý bude prehľadným spôsobom sprístupňovať a organizovať záznamy (a iné údaje) o korešpondencii, ktoré sme získali výskumom v rámci vyššie uvedeného projektu INDED.

Podmienky a návrh riešenia

Pri vypracovávaní návrhu realizácie predmetnej databázy sme vychádzali primárne z nasledovných podmienok:

- požiadavka minimalizácie nákladov na obstaranie riešenia,
- dostupnosť softvérových komponentov zdarma (najlepšie open source),
- maximálna jednoduchosť implementácie a používania systému,
- dlhodobá udržateľnosť riešenia,
- celková flexibilita a škálovateľnosť návrhu.

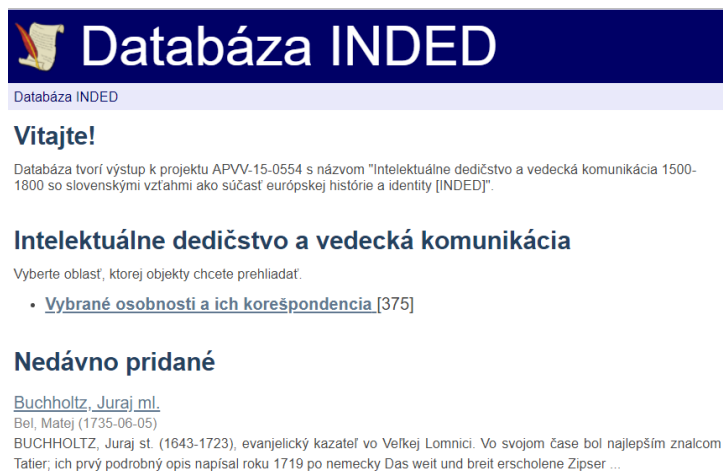
Primárnym cieľom štúdie bolo navrhnúť ucelené virtualizované riešenie, ktoré by spĺňalo uvedené požiadavky a súčasne poskytovalo funkčné prostredie vhodné pre umiestnenie a prevádzku databázy. Koncept štúdie prakticky kopíruje schému *open architektúry*[1] navrhutej špeciálne pre prípady použitia tohto typu, ktorú sme predstavili v rámci inej práce[2]. V zásade pokladáme nasledovné riešenie ako jednu z možných alternatív ku nepomerne drahším, zvyčajne komerčným riešeniam.

Základom celého návrhu je vhodný hardvér, ktorý by mal byť dostatočne výkonný. Z finančných prostriedkov uvoľnených pre projekt, sme obstarali lacnejší, no stále pomerne výkonný server osadený procesorom z rodiny Intel Xeon, ktorému sekundovalo 32GB operačnej pamäte a dostatok diskového priestoru v poli typu RAID. Použili sme bežne dostupné SATA disky so zvýšenou odolnosťou. Nad vrstvu hardvéru sme následne priamo nainštalovali zvolený virtualizačný hypervízor, ktorému sa budeme venovať podrobnejšie neskôr v rámci tohto článku. Predstavenie symbiózy virtualizovaného prostredia a systému DSpace spadá pod primárne ciele predloženého článku. Vo vybranom virtualizovanom prostredí sme vytvorili inštanciu virtuálneho stroja osadenú operačným systémom typu open source - Ubuntu Server s dlhodobou podporou. Zvolený operačný systém sa jednoducho udržiava, aktualizuje a spravuje, preto bola naša voľba v tomto smere pomerne jednoznačná. V prípade problémov sú dostupné viaceré odborné fóra udržiavané odbornými komunitami po celom svete.

Inštancia DSpace ako online databáza

Nad operačným systémom bol, ako súčasť aplikačnej vrstvy, inštalovaný systém DSpace [4], ktorý je dnes známy predovšetkým ako zdarma dostupný softvérový základ pre podporu digitálnych knižníc a open access repozitárov. Tentokrát sme sa však rozhodli ho použiť alternatívnym spôsobom – a síce v úlohe online databázy určenej na ukladanie a sprístupňovanie metadát viažucich sa ku v projekte skúmanému objemu vedeckej korešpondencie. Problematika možného sekundárneho využitia systémov tohto typu ďaleko presahuje možnosti predloženého článku, a preto sa jej budeme venovať v inej, samostatnej štúdií.

Konkrétne listové metaúdaje, skeny a registry, ktoré budú obsiahnuté v predmetnej online databáze, boli zozbierané počas mnohých praktických výskumov realizovaných vo viacerých pamäťových inštitúciách. Tieto dáta boli následne vhodným spôsobom koncentrované, naformátované do súborov typu CSV[3] a importované do prázdnej inštancie DSpace. Každý list predstavoval samostatný záznam s unikátnym ID a vopred definovanou metadátovou schémou odvodenou zo štandardu Dublin Core. Jednoduchý screenshot upraveného rozhrania systému možno vidieť na obrázku 1 nižšie.



Databáza INDED

Databáza INDED

Vitajte!

Databáza tvorí výstup k projektu APVV-15-0554 s názvom "Intelektuálne dedičstvo a vedecká komunikácia 1500-1800 so slovenskými vzťahmi ako súčasť európskej histórie a identity [INDED]".

Intelektuálne dedičstvo a vedecká komunikácia

Vyberte oblasť, ktorej objekty chcete prehliadať.

- [Vybrané osobnosti a ich korešpondencia](#) [375]

Nedávno pridané

[Buchholtz, Juraj ml.](#)
Bel, Matej (1735-06-05)
BUCHHOLTZ, Juraj st. (1643-1723), evanjelický kazateľ vo Veľkej Lomnici. Vo svojom čase bol najlepším znalcom Tatier, ich prvý podrobný opis napísal roku 1719 po nemecky Das weit und breit erscholene Zipser ...

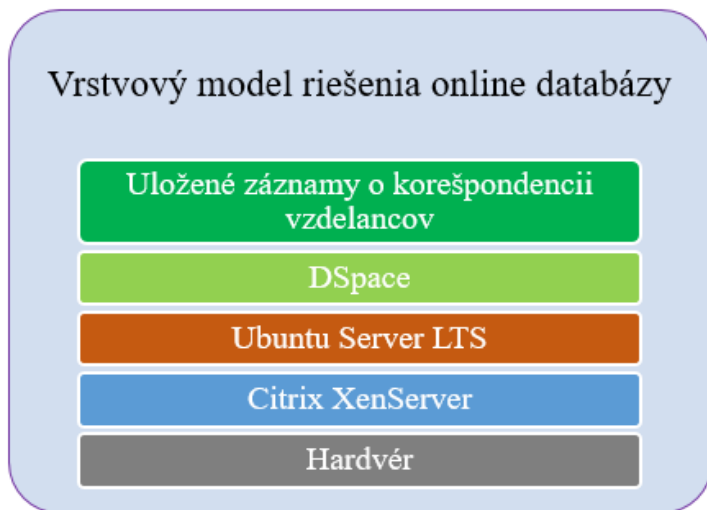
[5]

Rozhranie databázy INDED

Citrix XenServer

Ako uvádzajú autori Singh a Chu (2016), XenServer [6] vznikol ako výsledok výskumného projektu na Univerzite v Cambridge v roku 2003. Postupne získal presah aj do komerčnej sféry. V roku 2007 prešiel pod záštitu nadnárodnej spoločnosti Citrix, ktorá sa rozhodla tento produkt naďalej vyvíjať a poskytovať k základnej verzii zdarma i komerčne dostupné varianty spoločne s pokročilými funkcionalitami a platenou podporou. Turner, Barron a McCrocklin (2016) dodávajú, že cieľom spomínaného projektu je neustále poskytovať a udržiavať bezpečnú platformu umožňujúcu beh viacerých operačných systémov na jednom hostujúcom serveri.

Prechod open source projektu pod Citrix podnietil rozvoj komunity vývojárov, testerov i nadšencov, ktorí pomáhajú produkt naďalej udržiavať. I vďaka tomu si Citrix XenServer získal vo svojej bezplatnej verzii mnoho používateľov v menšom firemnom i akademickom sektore. Aj to boli rozhodujúce dôvody, pre ktoré sme nasadili riešenie spoločnosti Citrix v konkrétnom akademickom prostredí – na Katedre mediamatiky a kultúrneho dedičstva FHV Žilinskej univerzity v Žiline. Virtualizovaná online databáza vedeckej korešpondencie teda priamo využíva tento softvér, čo môžeme vidieť aj na priloženom obrázku 2 odrážajúcom navrhovaný koncept riešenia.



[7]

Vrstvový model riešenia

Kvôli zachovaniu objektivity je však nutné podotknúť, že spoločnosť Citrix sa rozhodla v súčasnosti klásť dôraz predovšetkým na presadzovanie pokročilých enterprise funkcionalít, čo malo za následok celkové razantné obmedzenie možností bezplatnej verzie XenServera. Základné funkcionality sú stále dostupné, no oklieštenia bezplatnej verzie, ktoré Citrix s vydaním verzie 7.3, spôsobili odštiepenie významnej časti vývojárskej komunity a vznik oddeleného open source projektu s názvom XCP-ng [8] (Pezin 2018), teda *Xen Cloud Platform - New Generation* (Lambert 2018).

Ako píše **Pezin** (2018) a **Lambert** (2018), nový projekt XCP-ng má byť zabezpečovaný výhradne komunitou bez účasti spoločnosti Citrix. Produkt má navyše disponovať všetkými pokročilými funkcionalitami, ktoré boli odobraté pôvodnej verzii XenServera v roku 2018. Situáciu ohľadom ďalšieho vývoja oboch produktov budeme i naďalej priebežne monitorovať. Neskôr plánujeme odbornej verejnosti priniesť i podrobnú komparatívnu štúdiu oboch ponúkaných riešení vychádzajúcu z praktického testovania v konkrétnom akademickom prostredí.

Záver

V predloženom článku sme sa snažili poukázať na skutočnosť, že virtualizácia serverov sa stáva vďaka viacerým voľne dostupným riešeniam dostupnou i pre organizácie, ktoré si nemôžu dovoliť financovať komplexné riešenia. Samozrejme, i implementácia open source produktu si vyžaduje nemalé úsilie a technické znalosti, no v konečnom dôsledku zvyčajne napomáha k významným úsporám, ktoré možno následne použiť iným spôsobom. Ukázali sme, že alternatívnym spôsobom možno využiť aj známy softvér DSpace – ktorý s prehľadom zvládne ukladať a sprístupňovať požadované metadáta o nesmierne variabilných entitách. V našej konkrétnej štúdiu sme predstavili použitie DSpace v úlohe online databázy sprístupňujúcej metaúdaje o vedeckej korešpondencii vzdelancov v Európe medzi rokmi 1500-1800.

Hoci databáza momentálne obsahuje len záznamy o korešpondencii **Mateja Bela**, je už online k dispozícii k prvotnému nahliadnutiu na adrese <http://inded.fhv.uniza.sk> [9]. Čoskoro budú do databázy pridané i záznamy o korešpondencii ďalších významných osobností. Technické a bezproblémové zabezpečenie prevádzky a zálohovania nám poskytuje momentálne virtualizačný hypervízor XenServer 7.6 vo verzii dostupnej zdarma. Či dôjde niekedy k zmene použitej platformy ukáže až čas, ďalší výskum, testovanie a analýza praktických skúseností.

Literatura:

- FORMANEK, M. 2018. Reálny a virtuálny verejný priestor [Dizertačná práca]. Žilinská univerzita v Žiline. Fakulta humanitných vied; Katedra mediamatiky a kultúrneho dedičstva. Školiteľ: prof. Ing. Milan Konvit, PhD. Žilina: Katedra mediamatiky a kultúrneho dedičstva.

- JOHNSTON, R. et al. Xen Network Flow Analysis for Intrusion Detection. IN : *CISRC '16 Proceedings of the 11th Annual Cyber and Information Security Research Conference*. 2016. ISBN 978-1-4503-3752-6 DOI: 10.1145/2897795.2897802
- LAMBERT, O. 2018. *What is XCP-ng?* [online]. 2018-11-28 [cit. 2019-04-04]. Dostupné na: <https://github.com/xcp-ng/xcp/wiki/FAQ> ^[10]
- PEZIN, Marc-André. 2018. *XCP-ng an open source alternative to XenServer* [online]. 2018-06-25 [cit. 2019-03-28] Dostupné na : [//medium.com/@marc.pezin](https://medium.com/@marc.pezin/xcp-ng-an-open-source-alternative-to-xenserver-29a7dc876af4) ^[11]; [// --> /xcp-ng-an-open-source-alternative-to-xenserver-29a7dc876af4">https:](https://medium.com/@marc.pezin/xcp-ng-an-open-source-alternative-to-xenserver-29a7dc876af4)