

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra počítačové grafiky a interakce



Zpráva k projektu 493/2013/1

Archivace dat s využitím DÚ CESNET

Jan Kubr, Ondřej Votava

Květen 2015

Úvod

Při řešení projektu bylo nutné navrhnout a implementovat pracoviště usnadňující archivaci dat na Katedře počítačové grafiky a interakce. Při návrhu pracoviště bylo nutné zohlednit připojení k datovým úložištím CESNET. Během řešení jsme specifikovali požadavky kladené na archivaci dat na našem pracovišti. Provedli jsme rešerši existujících řešení určených pro dlouhodobou archivaci dat a zvolené řešení jsme otestovali. Výsledkem je implementace archivačního pracoviště, které v současnosti funguje ve zkušebním provozu.

Současný stav řešeného problému

Dlouhodobá **archivace** dat se na našich pracovištích neprovádí.

Scénáře archivace dat

Pro archivaci dat navrhujeme následující scénáře členěné podle typu archivovaných dat. Scénáře a operace požadované ve scénářích definují výsledné požadavky na archivační systém.

Individuální soubory uživatele

Popis

Ukládám soubor pro vlastní potřebu. Nikdo jiný je nesmí vidět. Soubory mohou být zašifrované.

Operace

Uložení,
Smazání,
Smazání po expiraci.

Příklad: soukromé výzkumné zprávy, data z měření pro vlastní potřebu, archiv emailů.

Sdílené soubory

Popis

Ukládám soubory společného projektu. Data jsou sdílena v rámci skupiny uživatelů. Soubory mohou být zašifrované, ale musí je umět rozšifrovat celá skupina.

Operace

Uložení,
Smazání,
Smazání po expiraci,
Sdílení odkazu.

Příklad: výzkumné zprávy, data z měření.

Veřejné soubory

Popis

Ukládám veřejná data. Soubory nejsou zašifrované.

Operace

Uložení,
Smazání,
Smazání po expiraci,
Sdílení odkazu.

Příklad: veřejné výzkumné zprávy, veřejná data z měření.

Elektronicky podepsané dokumenty

Popis

Ukládám soukromá, sdílená či veřejná data. Soubory jsou elektronicky podepsané.

Operace

Uložení,
Přepodepsání,
Smazání,
Smazání po expiraci,
Smazání po expiraci podpisu,
Sdílení odkazu.

Příklad: podepsaná závěrečná zpráva, podepsaná přihláška projektu, podepsaný dokument typu bezpečnost práce.

Popis operací

Uložení

1) Vytvořím soubor, otaguji jej, informace uložím do DB, (zašifruji), uložím soubor na úložiště.
2) Vytvořím soubor, (zašifruji), uložím soubor na úložiště, systém jej otaguje, informace systém uloží do DB, připojím poznámky k souboru.
pozn.: Nemožnost otagování u zašifrovaného souboru.

Vyhledání

Vyhledání souboru podle:

času

vzniku
modifikace
přístupu
expirace

jména

přípony

typu (tagu)

frekvence přístupů

Smazání

Smazání podle vyhledání (filtru)

Automatické smazání po expiraci
s potvrzením
bez potvrzení.

Automatické smazání po expiraci podpisu
s potvrzením
bez potvrzení.

Sdílení odkazu

Vytvoření URL pro sdílení souboru.

Přepodepsání

Vytvoření nového podpisu pro zamezení vypršení platnosti podpisu expirací původního certifikátu.

Existující řešení

Piler (<http://www.mailpiler.org/en/index.html>)

Systém pro archivaci emailů, nabízí tvorbu skupin, které mohou sdílet emaily, politiky pro archivaci, např. jak dlouho má být jaký typ emailu (podle odesilatele, příjemce atd.) archivován a následně smazán z archivu. Pošta může být prohledávána a tříděna dle mnoha kritérií. Piler nabízí i specifický režim pro auditory, kdy auditor může číst všechnu poštu.

Metadata jsou ukládána do databáze (MySQL), emaily jsou uloženy jako soubory, přičemž všechny soubory jsou šifrovány. Přístup k emailům zajišťuje webové rozhraní. Piler lze propojit s mnoha emailovými servery (všechna pošta je přeposílána na Piler), lze jej integrovat s Active Directory pro Single Sign On. Autentizaci uživatelů může zajišťovat také LDAP či IMAP.

Shrnutí

Piler dokáže pokrýt většinu našich požadavků na archivaci jako vyhledávání, třídění, automatické mazání či sdílení. Je však úzce zaměřen pouze na jedinou oblast emailů. Jeho nasazení navíc vyžaduje změny v infrastruktuře celé organizace (je třeba upravit emailový server). Proto systém Piler nevyhovuje našim požadavkům jako systém pro archivaci dat.

OpenKM (<http://www.openkm.com/en/>)

Pokročilý nástroj pro správu dokumentů. Disponuje webovým java rozhraním pro správu dokumentů. OpenKM podporuje následující pokročilé funkce (P komerční Professional verze):

hromadné přidávání dokumentů (P),

opatření dokumentů metadaty,

digitální podpisování dokumentů (P),

kontrola expirace dokumentu (P),

vyhledávání,

verzování dokumentů.

Celá řada vlastností je uvedena na stránkách projektu v porovnání jednotlivých verzí

(<http://www.openkm.com/en/product/comparison-of-versions.html>).

Požadavky:

Server

Java 1.7 a vyšší

Apache Tomcat

Operační systémy: Unix, GNU/Linux a Windows

Data jsou ukládána na souborovém systému, nebo DBMS (Oracle, PostgreSQL, MySQL, MS SQL Server, ...).

Shrnutí

Jedná se o pokročilý archivační nástroj, který splňuje naše požadavky v komerční verzi. Pro bezproblémové použití je potřeba použít tuto komerční verzi, nebo naprogramovat rozšíření pro verzi komunitní.

LogicalDOC (<http://www.logicaldoc.com/>)

Správce souborů a informací LogicalDOC nabízí kompletní systém pro správu dokumentů, včetně indexace, verzování, prohledávání či sdílení. Další funkce, jako např. náhledy v PDF, podepisování dokumentů či online editace nabízí pouze v placené verzi.

Data lze do aplikace nahrávat přes webové rozhraní, díky množství API lze využít i desktopové či mobilní klienty. LogicalDOC nabízí jako jeden modul archivaci (pouze v placené verzi). Pojetí archivace je však odlišné od našeho. Je-li soubor archivován, pak k němu nemá přístup nikdo jiný, než správce, jeho obsah je přenesen na specifické úložiště, kde již soubor není indexován aby nezatěžoval systém při práci s nearchivními dokumenty.

Shrnutí

LogicalDOC je nástroj s mnoha funkcemi. Pro archivaci není vhodný, neboť jeho přístup k archivovaným souborům je zcela odlišný, než požadujeme. Pokud bychom nevyužívali funkci archivace a všechna data měli v aktivním režimu, pak systém LogicalDOC neposkytuje nástroje pro omezení životnosti dokumentu.

OpenDocMan (<http://www.opendocman.com>)

Opensource systém pro správu souborů. Nabízí standardní řešení pro limitování přístupu k souborům (jak na úrovni uživatelů, tak uživatelských skupin), jejich indexaci (pouze metadata vyplněná uživatelem), historii verzí, zamykání souboru proti změnám či schvalování nahraných souborů správcem. OpenDocMan nabízí ke stažení plug-in, který umožňuje nastavení expirace souboru (každý uživatel si může pro své soubory nastavit expiraci).

Nevýhodou aplikace OpenDocMan je její uživatelské rozhraní, které je málo přehledné a pro mnoho činností zbytečně komplikované. Správa uživatelů je čistě v režii aplikace, nelze ji napojit na běžné adresářové služby.

Epiware Document Management (<http://www.epiware.com>)

Další systém pro správu dokumentů. Nabízí historii revizí, zamykání dokumentu proti změnám, řízení přístupu k souborů podle skupin či uživatelů. Aplikace je k dispozici pouze pro OS Windows.

Zvolené řešení

Pro testování jsme použili nástroj OpenKM.

Popis Instalace

Ze stránek výrobce (<http://www.openkm.com/>) lze stáhnout balík obsahující vše potřebné pro chod aplikace. Po rozbalení balíku lze aplikaci OpenKM rovnou spustit (v našem případě balík pro OS Linux obsahoval jak běhové prostředí javy, tak nakonfigurovaný aplikační server Tomcat). Pro spuštění aplikace pak stačilo spustit Tomcat a případně vytvořit spouštěcí skript pro OS (aby aplikace běžela automaticky po startu OS).

Následně byla na serveru vytvořena složka, která je připojena přes NFSv4 k úložišti Cesnet. Aplikace OpenKM byla nastavena tak, aby data ukládala do této složky (úprava konfiguračního skriptu dle [manuálu](#)). Po restartu aplikace byla již všechna data ukládána na DU Cesnet.

Pro běh aplikace OpenKM jsou vyžadovány alepoň 2GB paměti RAM, po spuštění na našem serveru aplikace zabrala 1,5 GB, po delší době se hodnota zabrané paměti ustálila na 1,3 GB.

Výsledky testování

V rámci testování bylo provedeno základní testování funkcionality systému, které ukázalo řadu nedostatků po výchozí instalaci OpenKM. Mezi nejzávažnější nedostatky patřilo:

- chyba v certifikátu Java appletu, kdy není možné spustit editor a uploader v security režimu Javy, opraveno přidáním výjimek do konfigurace JRE,
- chyba v uploaderu (... IOException: ... (Operation not permitted)1700014975), nefunguje v Safari Verze 8.0.6 (10600.6.3), ve Firefox 37.0.2 funguje,
- chybějící možnost nastavení expirace souborů.

V rámci testování jsme provedli i základní výkonnostní testy:

1. vložení souboru ze vzdálené destinace,
2. vložení velkého souboru.

Vložení souboru ze vzdálené destinace

V průběhu testu jsme vkládali fotografie z hotelové sítě v Marrakéši. Síť měla v průběhu měření parametry uvedené v Tabulka 1. Výsledné časy pro upload fotografií jsou uvedeny v Tabulka 2.

Čas	Download (Mbit/s)	Upload (Mbit/s)	Odezva (ms)
21. 5. 2015 14:39	4,39	3,49	82
21. 5. 2015 14:37	4,25	1,02	80
21. 5. 2015 14:36	3,55	1,44	498
21. 5. 2015 14:14	7,07	1,23	85

Tabulka 1: Parametry sítě

Čas	Velikost (MB)	Doba pro upload (s)	Rychlost uploadu (Mbit/s)
21. 5. 2015 14:30	2,2	75	0,23
21. 5. 2015 14:32	2,6	46	0,45
21. 5. 2015 14:33	2,5	40	0,5
21. 5. 2015 14:40	2,4	30	0,64
21. 5. 2015 14:42	2,5	53	0,38

Tabulka 2: Rychlost vložení souboru

Vložení velkého souboru

Otestovali jsme možnost vložení velkého souboru v rámci školní sítě s následujícími výsledky:

Velikost souboru	Doba Uploadu (s)	Doba zobrazení v OpenKM (s)	Poznámka
650 MB	40	70	ISO soubor
3,7 GB	42	126	ISO soubor
3,7 GB	70	80 (zkopírování na disk)	srovnání v OpenNebule, upload přes HTTPS

Tabulka 3: Doba potřebná k vložení velkého souboru

Tabulka 3 ukazuje, jaký vliv má uložení dat na DU Cesnet na chování OpenKM. V řádku 3 je pro srovnání uvedena doba potřebná pro uložení stejného souboru jako na druhém řádku přímo na disk virtuálního stroje. Samotné kopírování na lokální disk trvá cca. 10 vteřin, při ukládání na DU Cesnet je ukládání výrazně delší, nikoliv však závažné.

Závěr

V rámci projektu byl zvolen nástroj OpenKM pro archivaci dokumentů na našem pracovišti. Z výsledku testů je zřejmé, že nástroj je možné používat i na připojeném DÚ CESNET.

Bohužel jsou pokročilé archivační funkce dostupné pouze v komerční verzi. V současnosti zvažujeme rozšíření komunitní verze o požadované funkce v rámci studentských projektů.