

FileNet Panagon → OpenText TCP Migrační report

Máme pro vás  řešení...

IXTENT s.r.o.
Jankovcova 2c
CZ - 170 00 Praha 7
Phone: +420 234 379 330
Fax: +420 234 379 331
sales@ixtent.com, <http://www.ixtent.com>

Impressum

Autor: Petr Olšovský

E-mail: petr.olsovsky@ixtent.com

Datum: 11.4.2012

Verze: 01

Copyright ©2019 IXTENT s.r.o.

Všechna práva vztahující se k tomuto dokumentu jsou vyhrazena. Žádná část tohoto dokumentu nesmí být reprodukována, předávána třetím stranám, obnovována elektronickými systémy, kopírována, distribuována či používána při veřejných prezentacích, a to v jakékoliv podobě, bez písemného souhlasu IXTENT s.r.o. Společnost IXTENT s.r.o. si vyhrazuje právo na změny obsahu.

Trademarks

IXTENT® je registrovanou obchodní známkou společnosti IXTENT s.r.o. Jména ostatních produktů byla použita výhradně za účelem jejich identifikace a mohou být obchodními známkami jejich výrobců.

Obsah

1	Úvod	4
2	Popis procesu migrace	4
3	Souhrnné informace	4
4	Schéma procesu migrace	5
5	Export dokumentů z Panagonu	7
5.1	Export seznamu ID dokumentů	7
5.2	Příprava dávek	7
5.3	Export obsahů ve formátu FOB	7
5.4	Generování XML a nativních formátů	7
5.5	Monitoring exportu	8
5.5.1	Export dokumentů	9
5.5.2	preImport dokumentů	10
6	Import dokumentů do TCP	11
6.1	Kopírování vstupních souborů	11
6.2	Prepipeline skript	11
6.3	Pipeline skript	11
6.4	Monitoring importu	12
7	Kontrola kvality migrace	13
7.1	Databázové zpracování	13
7.2	FN2OT Validator	13
7.3	DBCompare	14
7.4	Výsledek kontroly	14
8	Přílohy	15

1 Úvod

Dokument popisuje nástroje použité k migraci dokumentů ze systému IBM FileNet Panagon do prostředí OpenText TCP ve společnosti Telefónica Slovakia. Je zde popsán průběh migrace se zaměřením na výjimečné stavy, problémy a řešení. A také je zde zaznamenána rychlost jednotlivých migračních kroků.

Dále dokument obsahuje informace o kontrole úspěšnosti migrace, použité nástroje a výsledek kontroly.

2 Popis procesu migrace

Migrace byla původně navržena jako dvoukrokový proces. V prvním kroku mělo dojít k úplnému exportu dokumentů a jejich metadat z Panagonu, a ve druhém k importu do TCP.

Prvotní testy a odhady časové náročnosti exportu ale vedly k přehodnocení návrhu, a rozdělení exportu na další dílčí kroky. Kromě toho byly provedeny podstatné změny, a sice:

- Byl opuštěn koncept vyhledávání dokumentů přímo exportní aplikací a byly použity off-line databázové dotazy a skripty.
- Pro řízení migrace byla použita SQL databáze, kde byla uchovávána data o jednotlivých dávkách i migrovaných dokumentech v nich.
- Počet dokumentů v pracovních adresářích byl snížen ze 100.000 na pouhých 10.000, čímž došlo k výraznému zrychlení diskových operací.
- Rychlost exportu byla dále zvyšována počtem použitých paralelních instancí jednotlivých nástrojů.

Popsané optimalizace vedly ke zvýšení výkonnosti a rychlosti migrace, na druhou stranu však podstatně zkomplikovaly průběh a vyžadovaly vyšší nároky na obsluhu.

Import dokumentů byl proveden použitím tzv. Document Pipelines (dále DP). Import byl prováděn v jedné instanci, další se nepodařilo zprovoznit.

3 Souhrnné informace

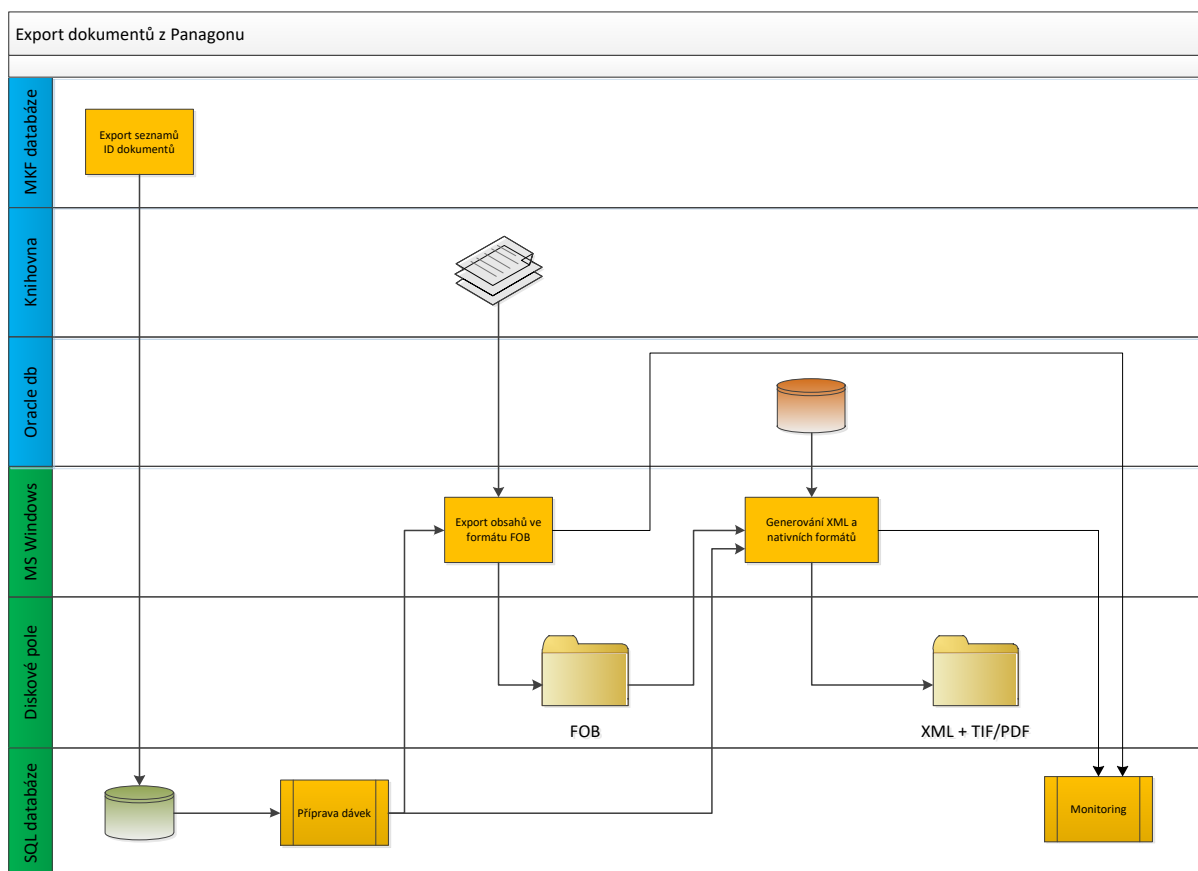
Dosud bylo provedeno několik migračních běhů tak, jak shrnuje následující tabulka:

Zahájení migrace	18.10.2011
Počet dokumentů v iniciálním běhu migrace	19 142 589
Dokončení iniciálního běhu migrace	4.1.2012
Počet dokončených přírůstkových migračních běhů	9
Dokončení posledního přírůstkového běhu migrace	11.4.2012
Celkový počet zmigrovaných dokumentů po posledním přírůstkovém běhu	23 570 112
- počet dokumentů - Kontrakt	2 013 998
- počet dokumentů - Bimage	9 611 860
- počet dokumentů - Dimage	11 944 244
Identifikátor posledního zmigrovaného dokumentu (F_DOCNUMBER)	27 210 430

4 Schéma procesu migrace

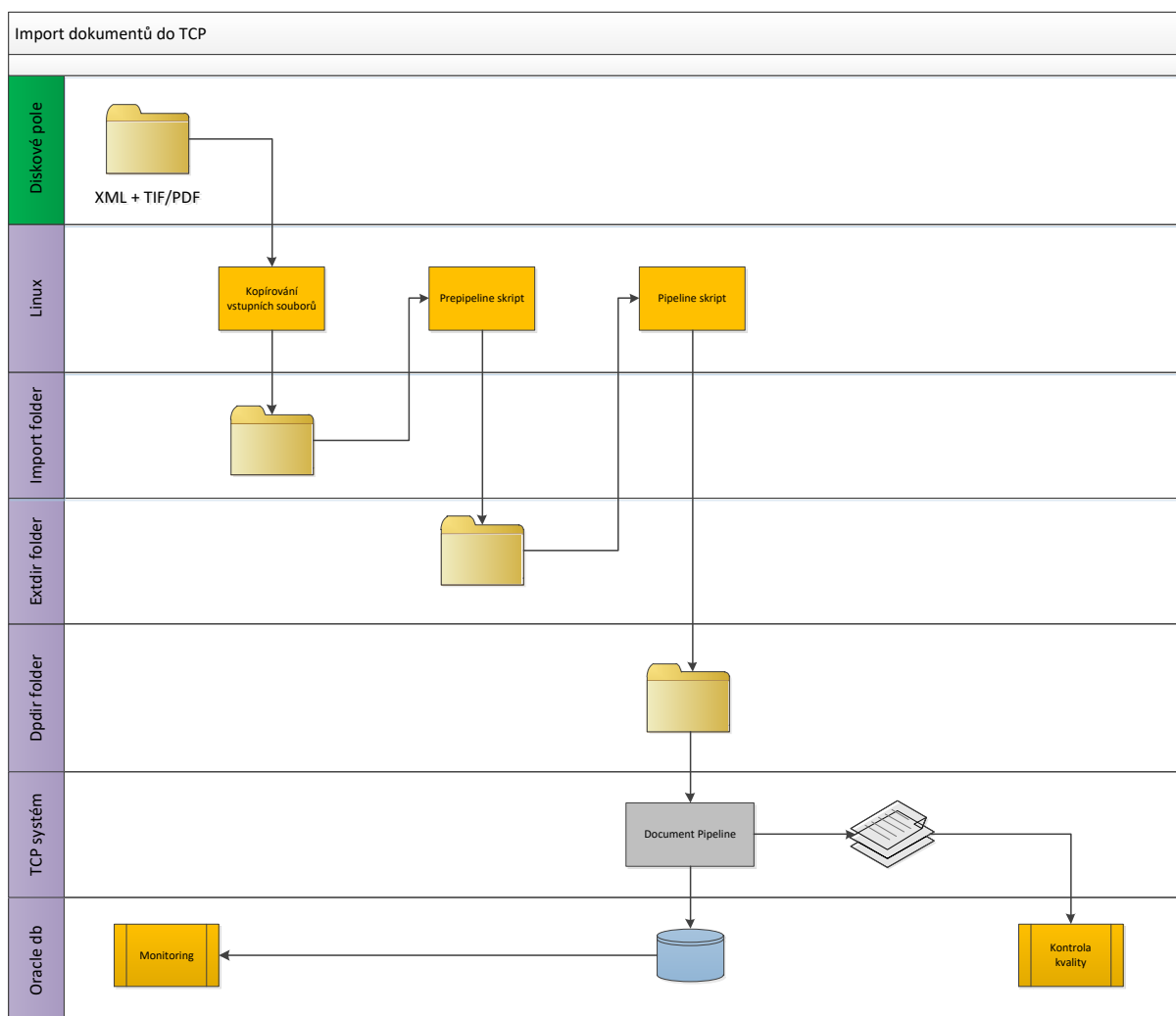
Export dokumentů z Panagonu se skládal z následujících kroků:

Krok	Podrobnější popis
Export seznamu ID dokumentů	Vytvoření seznamu dokumentů podle médií v MSAR/CSAR použitím MKF_tool.
Příprava dávek	Vytvoření optimálně velikých dávek pro další exportní nástroje s ohledem na počet instancí a zátěž Panagonu.
Export obsahů ve formátu FOB	Stažení obsahů dokumentů ve formátu FOB (FileNet Object) a uložení na diskové pole.
Generování XML a nativních formátů	Stažení metadat k dokumentům, zápis do XML souborů. Konverze FOB souborů do nativních formátů, spojování vícestránkových TIFF souborů do jednoho multipage TIFFu.
Monitoring	Evidence rychlosti zpracování dávek a počtu zpracovaných dokumentů. Centrální evidence chyb.



Import dokumentů do TCP se skládá z následujících kroků:

Krok	Podrobnější popis
Kopírování vstupních souborů	Kopírování dokumentů a indexových XML souborů na linuxový aplikační server OpenText TCP.
Prepipeline skript	Konverze vstupních dat do formátu požadovaného DP.
Pipeline skript	Load dokumentů do TCP prostřednictvím DP, mazání vstupních dat z aplikačního serveru
Monitoring	Evidence rychlosti zpracování dávek, počtu zpracovaných dokumentů a evidence chyb.
Kontrola kvality migrace	Kontrola úspěšnosti a kvality migrace.



5 Export dokumentů z Panagonu

5.1 Export seznamu ID dokumentů

Seznam dokumentů byl získáván dotazem z tabulky docs v utilitě MKF_tool.

Např. tento příkaz lze použít pro vytvoření seznamu dalšího přírůstkového migračního běhu:

```
s docs 27210430 < doc_id < 40000000 showing doc_id pages surface_id 1
```

5.2 Příprava dávek

Vyexportovaný seznam byl importován do MS SQL databáze PANAGON na serveru NTKOFAXS02, do tabulky dbo._Export_Source.

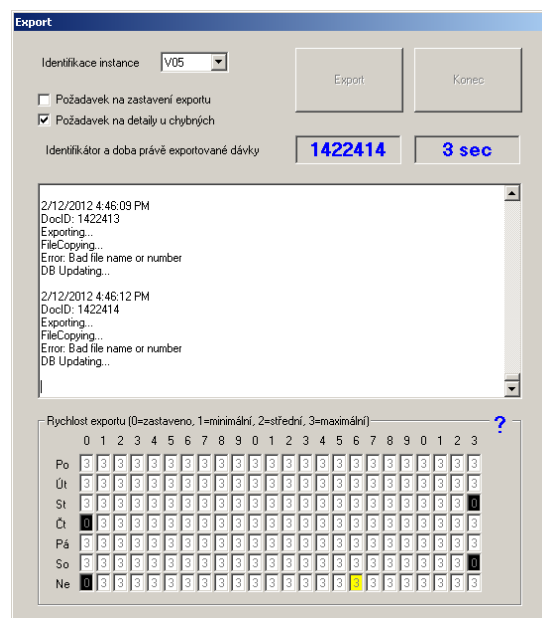
Pole „performer“ definuje instanci exportéru (aktualizováno ručně - updatem dle potřeby), pole „result“ pak nastavuje dokument ke zpracování utilitou Export nebo obsahuje výsledek zpracování, viz dále. Pro používané SQL příkazy viz Export.txt v přílohách.

5.3 Export obsahů ve formátu FOB

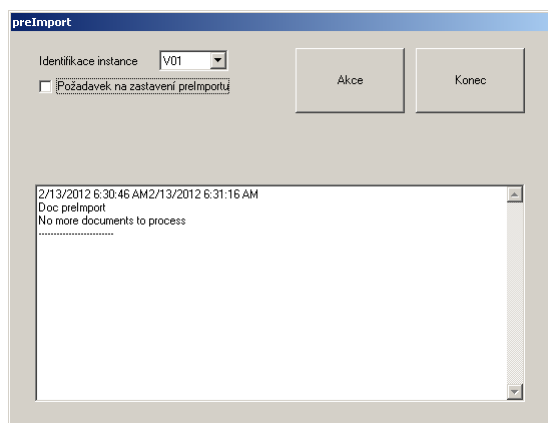
Označené dokumenty (pole „result“ s hodnotou NULL) byly exportovány z Panagonu v instancích (pole „performer“) nástrojem Export.

Tento nástroj vyvinutý ve VB 6.0 exportoval dokumenty z Panagonu ve formátu FOB (FileNet Object file) na diskové pole a aktualizoval pole „result“ (hodnota OK v případě úspěchů, ERR v případě nějaké chyby), „result_text“ (prázdné nebo chybová hláška) a „path“ (cesta k vyexportovanému obsahu).

Exportér fungoval po dávkách s průměrnou rychlostí 40 až 70 sekund na 100 exportovaných dokumentů. Byl spouštěn ve virtuálních serverech ve 3 až 5 paralelních instancích, vyšší počet již nebyl efektivní.



5.4 Generování XML a nativních formátů



Vyčítání metadat k vyexportovaným obsahům z Oracle databáze Panagonu a zápis do XML souborů na diskové pole bylo prováděno nástrojem preImport.

Tento nástroj z VB 6.0 dále prováděl konverzi obsahu z FOB na nativní formát a v případě vícestránkových TIFF dokumentů ještě vytváření multipage TIFF souborů.

Pro zvýšení výkonnosti byla obsluha tohoto nástroje ovládána tabulkou dbo._Export_Source_DONE, kam byly záznamy úspěšně vyexportovaných dokumentů („result“ s hodnotou OK) přesouvány

z tabulky `dbo._Export_Source` (obě tabulky v SQL databázi PANAGON na serveru NTKOFAXS02). Pro používané SQL příkazy viz `Export.txt` v přílohách.

Po dokončení `prelImportu` byla aktualizována pole „process“ (hodnota OK v případě úspěchu, CLASS pro dokumentovou třídu, která nebyla určena k migraci a ERR v případě nějaké chyby), „extension“ (s hodnotou nejčastěji PDF nebo TIF, ale i jinými) a „class“ (na hodnotu `F_DOCCLASSNAME` z Panagonu, tzn. 1 pro Kontrakt, 6 pro Korespondenci a 7 pro Bimage).

`prelImportér` byl spouštěn nejčastěji ve 3 instancích, každá z nich byla schopna v průměru zpracovat 1000 dokumentů za 150 až 300 sekund.

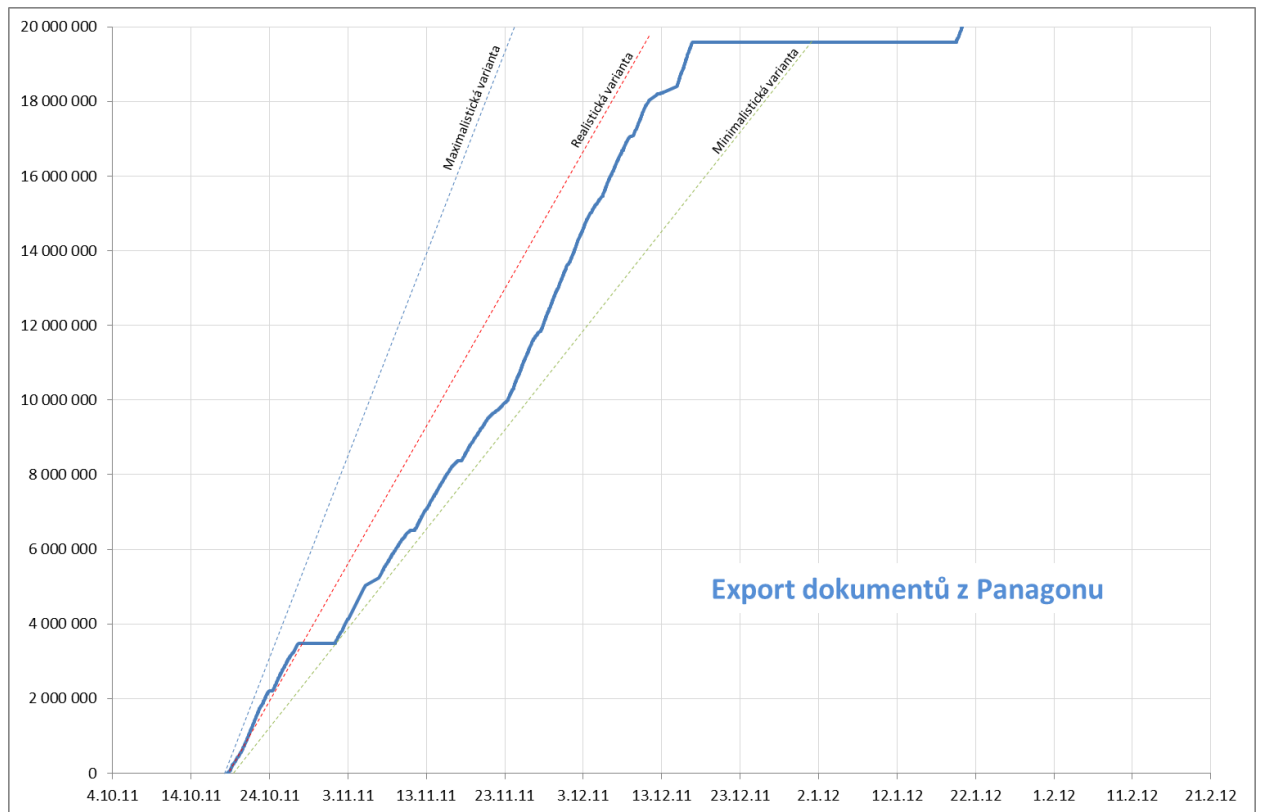
5.5 Monitoring exportu

Oba nástroje, `Exporter` i `prelImporter`, zaznamenávaly počet zpracovaných dokumentů do tabulky `dbo._Export_Statistics` v SQL databázi PANAGON na serveru NTKOFAXS02. K zápisu byly použity uložené procedury `Export_GenerateStatistics` a `PrelImport_GenerateStatistics`.

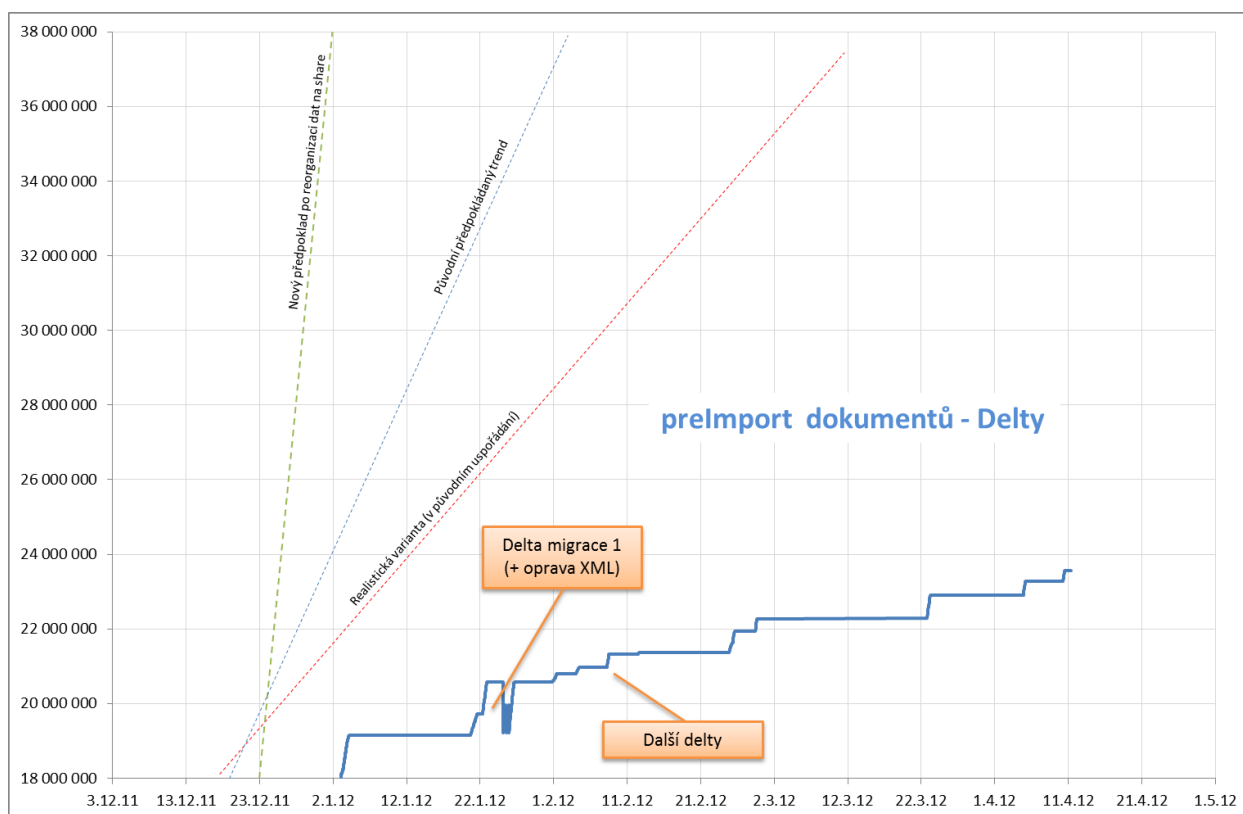
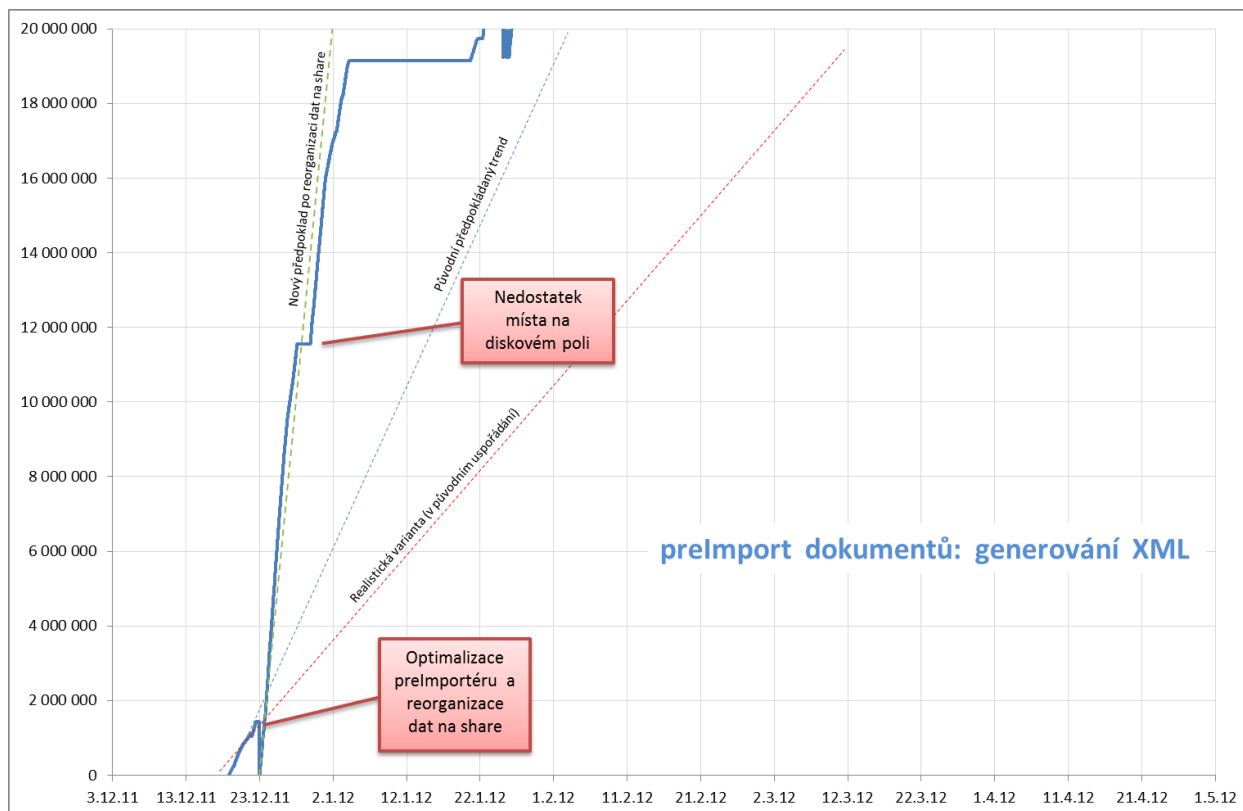
Ze zaznamenaných dat byly vytvořeny následující grafy. Další informace a podkladová data jsou k dispozici v příložených souborech `Export.xlsx` a `prelImport.xlsx`

Při exportu nedocházelo k žádným kritickým chybám. Očekávané chyby s infrastrukturou (nedostupnost diskového pole nebo databáze) byly řešeny opakováním chybných operací. Chyby s neexistencí dokumentů v Panagonu byly později řešeny ručně, viz kapitola `Kontrola kvality migrace`.

5.5.1 Export dokumentů



5.5.2 preimport dokumentů



6 Import dokumentů do TCP

6.1 Kopírování vstupních souborů

Vyexportované dokumenty s popisnými XML soubory byly ukládány exportními nástroji do jednoúrovňové stromové struktury adresářů (název adresáře byl sestaven ze čtyř levých znaků identifikátoru dokumentu).

Diskové pole bylo hostováno linuxovým serverem, což významně zjednodušilo kopírování na aplikační server OpenText TCP.

Kopírování bylo prováděno příkazem scp do předem připraveného adresáře (na obou stranách byly uloženy veřejné klíče), např. pro dávku dokumentů s identifikátory 2560#### měly příkazy tuto podobu:

```
cd /otx_as/var/dp/extdir/import_migration/  
  
mkdir 2560  
  
scp -B -i /home/otadm/.ssh/id_dsa admin@10.101.33.134:/share/MD0_DATA/mig2/_IMPORT/2560/* 2560
```

Kopírování souborů na aplikační server bylo nejpomalejší částí a určovalo rychlost importu. Bez dopadu na rychlost migračního procesu bylo ale možné kopírovat nanejvýš ve dvou ssh oknech současně.

Kopírování příkazem scp bylo „automatizováno“ vytvářením jednoúčelových dávkových skriptů.

6.2 Prepipeline skript

Konverzi nakopírovaných dat do formátu akceptovatelného DP zajišťovala sada prepipeline skriptů, viz příložený soubor pipeline.zip

Skripty byly spouštěny automaticky z crontabu v intervalu 20 minut, nebo je bylo v případě potřeby možné spustit ručně příkazem:

```
/otx_as/DP_TCP/perl/prepipeline/job_prepipeline.sh
```

Skripty zpracovávaly jen zcela dokopírované dávky dokumentů, tento stav byl „nastavován“ vytvářením semafor-souboru LOG příkazem (např. opět pro dávku dokumentů s identifikátory 2560####):

```
touch /otx_as/var/dp/extdir/import_migration/2560/LOG
```

6.3 Pipeline skript

Enqueue proces byl spouštěn, a tím i load do OpenText TCP, pipeline skriptem, viz příložený soubor pipeline.zip

Skript byl spouštěn automaticky z crontabu v intervalu 20 minut, minutu před prepipeline skriptem, nebo jej bylo v případě potřeby možné spustit ručně příkazem:

```
/otx_as/DP_TCP/perl/prepipeline/job_pipeline.sh
```

6.4 Monitoring importu

Průběh importu do TCP nebylo možné jednoduše automatizovat, počty naimportovaných dat se tak musely zjišťovat občas ručně dotazem z BOL databáze TCP (první poskytuje celkový počet, druhý potom počty podle jednotlivých adresářů z exportérů):

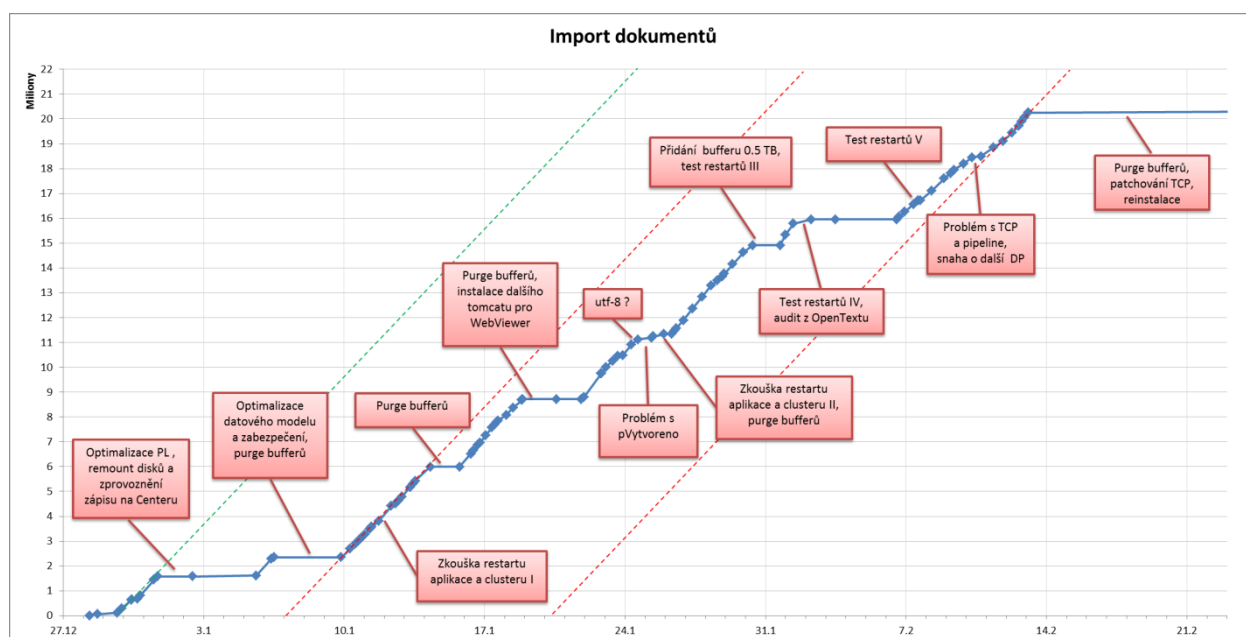
```
-- celkovy pocet naimportovanych
select count(*) from te2fa;

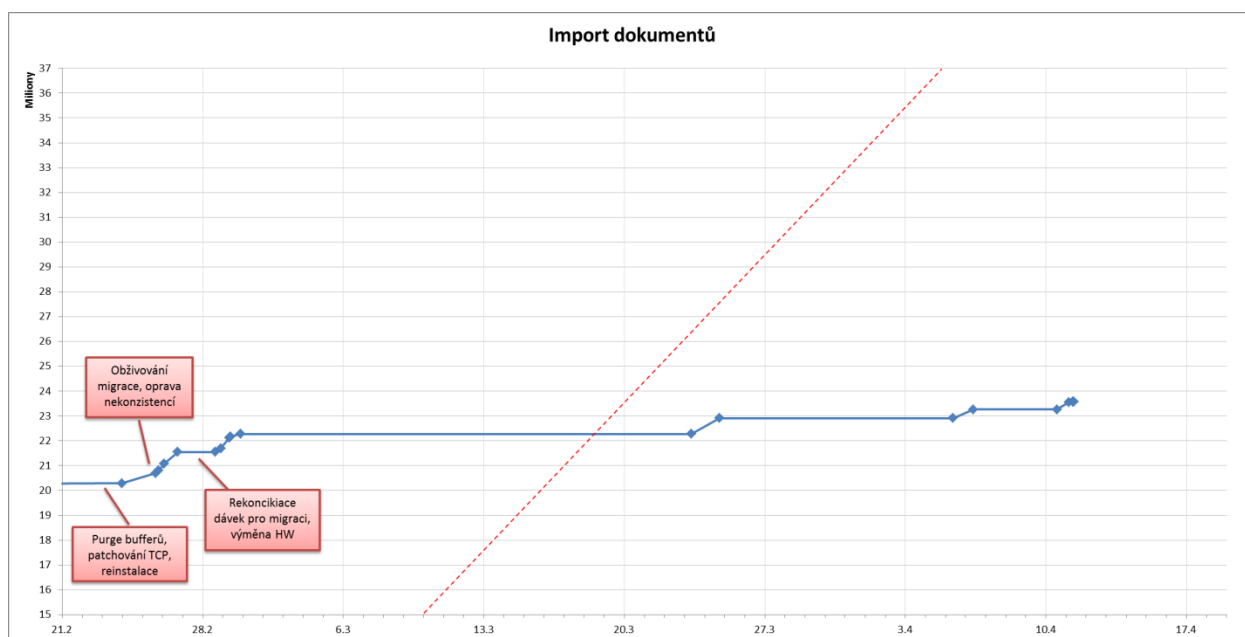
-- pocety vseh naimportovanych dokumentu podle Panagon_DocID folderu
select substr('00' || c111, length(c111)-5,4) folder, count(*) docs from te2fa
group by substr('00' || c111, length(c111)-5,4)
order by 1;
```

Při importu nastávaly problémy, které měly kritický dopad na průběh migrace. Všechny se ale nakonec, ať svépomocí či díky OpenText Supportu, podařilo vyřešit. Byly to zejména:

- Nefunkční platforma kvůli nastavení firewallů a dalších infrastrukturních záležitostí
- Nefunkční některá z komponent TCP, např. UMS
- Neustálé patchování/ instalace systémů
- Pády tomcatu a jBossu na aplikačním serveru
- Pomalé purgování bufferů
- Zaplnění disků (buffery, vstupní cache, disky s logy)
- Testování restartů clusteru

Průběh importu i problémy jsou částečně zachyceny v grafech, data jsou k dispozici v příloze Import.xlsx





7 Kontrola kvality migrace

Ke kontrole úspěšnosti migrace bylo použito těchto metod:

Metoda	Podrobnější popis
Databázové zpracování	Kontrola počtu dokumentů, zabezpečení a kritických vlastností databázovými prostředky s případnou opravou (Oracle PL SQL)
Nástroj FN2OT Validator	Namátková vizuální kontrola dokumentů a vlastností použitím jednoúčelového nástroje Validátor migrace (MS Visual Basic)
Nástroj DBCompare	Automatizovaná kontrola všech vlastností dokumentů s jejich případnou opravou (MS vb.net 2008)

7.1 Databázové zpracování

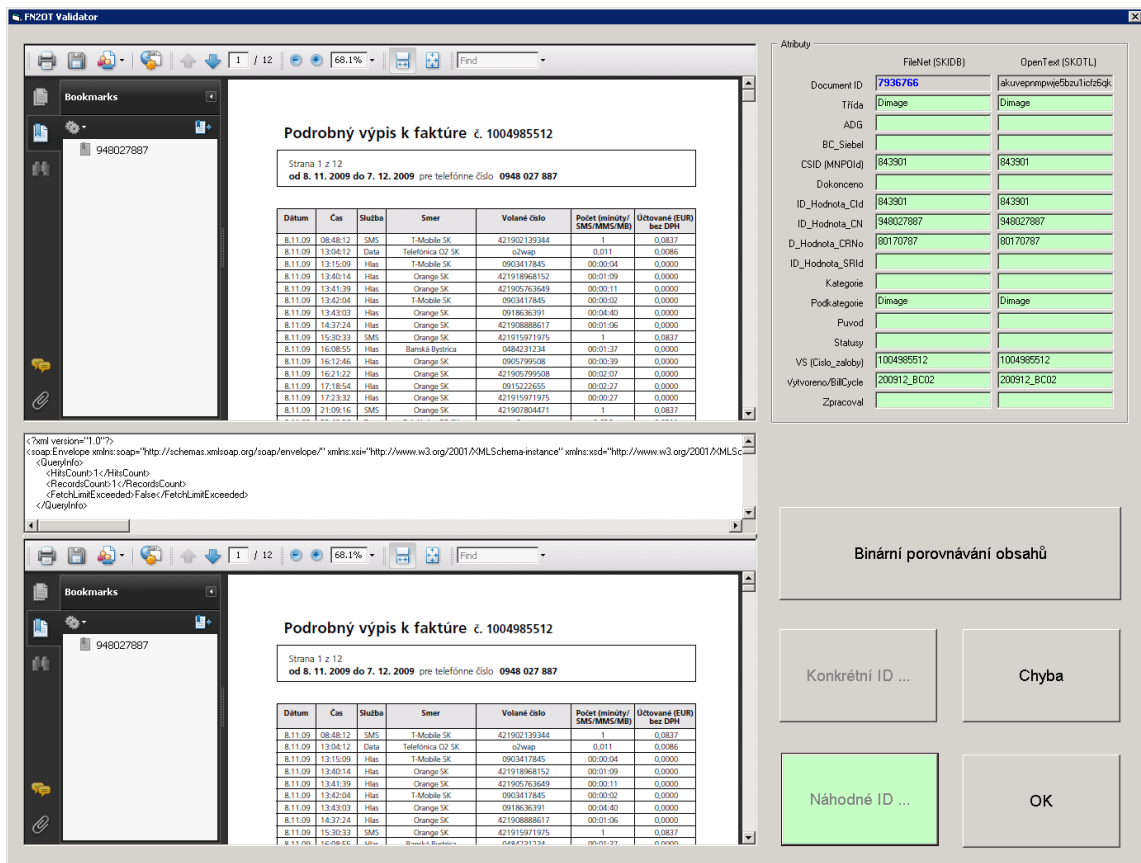
K porovnávání bylo použito databázového linku mezi Oracle databázemi Panagonu a TCP. Kontrola počtu dokumentů byla prováděna jednoduchými porovnávacími dotazy, jejichž výsledkem byl seznam identifikátorů dosud nezmigrovaných dokumentů, např.:

```
select d.f_docnumber, d.f_docclassnumber from doctaba@SKIDB_LNK d
where d.f_docnumber not in (select c111 from te2fa where
c10f in ('0Kontrakt','0Bimage','0Dimage'))
and d.f_docclassnumber in (1,6,7)
and d.f_docnumber <= 27210430;
```

Dále byly kontrolovány klíčové vlastnosti a byly prováděny aktualizace jejich hodnot. Dotazy a skripty jsou v příloze oracle.zip

7.2 FN2OT Validator

Jednoúčelová utilita pro vizuální porovnávání dokumentů a jejich vlastností v Panagonu a TCP je popsána v příloženém dokumentu Validátor migrace, v2.doc



Podrobný výpis k faktúre č. 1004985512

Strana 1 z 12
od 8. 11. 2009 do 7. 12. 2009 pre telefónne číslo 0948 027 887

Dátum	Čas	Služba	Směr	Volané číslo	Počet (minuty/ SMS/MMS/MB)	Účtovaná (EUR) bez DPH
8.11.09	08:48:12	SMS	T-Mobile SK	421902139344	1	0,0837
8.11.09	13:04:12	Data	Telefónica CZ SK	000000000000	0,011	0,0096
8.11.09	13:15:09	Hlas	T-Mobile SK	0903417845	00:00:04	0,0000
8.11.09	13:40:14	Hlas	Orange SK	421918968152	00:01:09	0,0000
8.11.09	13:41:39	Hlas	Orange SK	421905763649	00:00:11	0,0000
8.11.09	13:42:04	Hlas	T-Mobile SK	0903417845	00:00:02	0,0000
8.11.09	13:43:03	Hlas	Orange SK	0918636391	00:04:40	0,0000
8.11.09	14:37:24	Hlas	Orange SK	42190888617	00:01:06	0,0000
8.11.09	15:30:33	SMS	Orange SK	421915919175	1	0,0837
8.11.09	16:08:55	Hlas	Banka Bystřice	0484231234	00:01:37	0,0000
8.11.09	16:12:46	Hlas	Orange SK	0905799508	00:00:39	0,0000
8.11.09	16:21:22	Hlas	Orange SK	421905799508	00:02:07	0,0000
8.11.09	17:18:54	Hlas	Orange SK	0915222655	00:02:27	0,0000
8.11.09	17:23:32	Hlas	Orange SK	421915919175	00:00:27	0,0000
8.11.09	21:09:16	SMS	Orange SK	421907804471	1	0,0837

Tímto nástrojem bylo porovnáno 4 930 dokumentů, u 99 z nich byl hlášen problém (2%).
Detaily viz list Validace migrace v přiloženém souboru Import.xlsx

7.3 DBCompare

Utilita prochází definovanou sadu dokumentů zvolené dokumentové třídy a kontroluje veškeré dostupné atributy. Pouští se v maximálním možném počtu paralelních instancí.

Porovnává hodnoty atributů v Oracle databázích Panagonu a TCP, v případě rozdílu generuje databázové updaty, které se potom spouštějí ručně v SQL klientovi.

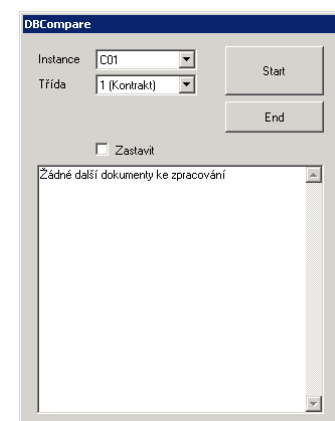
Sady dávek se definují v tabulce _Export_Source_DONE.

7.4 Výsledek kontroly

Z celého migračního procesu vypadlo přes 700 dokumentů, které se z různých důvodů nenamigrovaly. Důvodem byla nejčastěji nekorektní data (odporovala nastavení atributů v TCP) nebo problém s nesprávným mime-typem obsahů. Tyto dokumenty byly domigrovány jednotlivě po ručních opravách příčin a iterativním způsobem.

I přes maximální snahu se zatím nepodařilo přemigrovat 18 dokumentů s dále vypsány Panagon F_DOCNUMBER. Důvodem je nespíš to, že při pořizování těchto dokumentů v Panagonu neproběhla zcela operace COMMIT, data jsou v tabulkách doctaba a docs nekonzistentní.

1788351, 2537572, 10498797, 22102773, 22807441, 23103071, 23441088, 23447877, 23740008, 24041144, 24047733, 24263836, 25216587, 25218164, 25327827, 26992847, 27099189, 27193607



Podobné platí i pro následujících 12 dokumentů, kde je zase problém s obsahem dokumentů, kvůli jejich mime-typům a nebo proto, že jsou poškozené. Tyto dokumenty se nedaří exportními utilitami korektně vyexportovat.

727459, 734304, 734305, 804623, 3374594, 4910129, 5272030, 10498797, 17239791, 21830334, 23782268, 25851239

Ve všech těchto případech bude migrace pokud možno dokončena použitím TCP UI a manuální indexací.

Nejčastěji řešenými problémy v metadatech byly:

- diakritika v atributu Podkategorie z důvodu „nesprávného“ kódování exportních XML souborů počátečních dávek
- nesprávně naplněná hodnota atributu Kategorie u některých dokumentů třídy Kontrakt z důvodu pozměněného datového modelu a číselníků TCP oproti Panagonu
- nevyplněné CSID u dokumentů, kde obsažená hodnota MNPOld byla přesunuta do vlastního pole a CuRefNo zase do pole IDHodnota_CRNo (což byl jen nešvar z minulosti)
- chybné datumové a časové údaje v attributech Vytvoreno a Dokonceno (prohozený měsíc a den, případně posun o +/-1 hodinu z důvodu časů GMT/SELČ/SEČ)

Většina uvedených chyb byla opravena v Databázovém zpracování, oprava datumových položek ještě probíhá použitím nástroje DBCompare (nebylo snadné napsat db skript, proto probíhá s „inteligentnější“ aplikační logikou).

8 Přílohy

