Diplomová práce

Synchronizace a replikace geodat v prostředí Esri platformy

```
Úvod
1 CÍLE PRÁCE
2 POUŽITÉ METODY A POSTUPY PRÁCE
3 TEORETICKÁ VÝCHODISKA
   3.1 Použité programové prostředky
      3.1.1 SQL Server Express 2008
      3.1.2 PostgreSQL 9.x (PostGIS)
      3.1.3 ArcSDE geodatabase
   3.2 Použité postupy
      3.2.1 Replikace
      3.2.2 Synchronizace
      3.2.3 Verzování
   3.3 Geodata
4 VÝSLEDKY
5 DISKUZE
6 ZÁVĚR
Zkratky
ZDROJE
```

Úvod

1 CÍLE PRÁCE

Cílem diplomové práce je provést rešerši a na jejím základě prakticky otestovat proces synchronizace a replikace geodat, které se dnes objevují napříč platformou Esri. V teoretické části práce bude detailně analyzován proces synchronizace a replikace ve všech možných variantách (jednosměrná, dvousměrná, synchronní, asynchronní, ...) a popsány prostředky, které se na platformě Esri k těmto procesům využívají. Rozbor zahrne celé portfólio produktů od desktop řešení, přes možnosti ArcGIS serveru až po cloudový ArcGIS online. Budou popsány možnosti, požadavky a předpoklady pro úspěšnou realizaci.

V praktické části, nad existujícími katedrálními daty, dojde k praktickému testování těchto procesů na předem připraveném testovacím prostředí. Postupnými opakovanými procesy budou sledovány dílčí parametry procesu (rychlost procesu, úplnost, chybovost, podporované formáty). Vyjde se z primárně podporovaného databázového stroje SQL Server, který bude konfrontován s možnosti dalšího podporovaného systému PostgeSQL.

Můj jeden odstaveček - něco jako - jak vidím vlastní přínos do tématu.

2 POUŽITÉ METODY A POSTUPY PRÁCE

Napsat úplně nakonec, v rozsahu 1 stránky. Jaká data, jak probíhal postup.

NEVÍM KAM

- Produkty Esri nejsou databázové a pro uložení používají jiné databáze.
- Jakými kritérii se hodnotí, zda replikace probíhá správně it-literatura.
- Replikace rastrů je to možné? Zkoušel to už někdo?
- 4 úrovně použití replikace:
 - 1. uživatel ji nepotřebuje
 - 2. uživatel ji využívá skrze desktop (ale stále je zapotřebí server licence)
 - 3. přistupuje se síťově
 - 4. přistupuje se online
- <u>co budu sledovat</u> jestli databáze (databáze + SDE) umí replikovat jednoduchá i složitější geodata, jestli se to přeneslo, rychlost (Postgre x SQL), jaká bude kvalita (např. dle zachovaného tvaru, stejného počtu bodů, objemu), verifikace jestli vrátí info o tom, že to proběhlo
- zdroje: OGC, knihy, BostonGIS.com, dbsvět.cz (Co umí server 2008 česky), novinky 2012 (opravy, co je nového)
- Esri scí vektor, raster, pro uložení v databázi feature datasete (ukázka)
- rozhraní pro Postgre, SQL -- jak se ovládá ArcCatalog, v čem se nastavuje ArcSDE, SQL Tools "ArcCatalog pomáhá organizovat a spravovat veškerá data GIS, tj. mapy, glóby, dataové sady, geodatabáze, modely, metadata a služby (Esri, 2006)."
- 2.2.1? podkapitoly kde konfigurák, jak, výseky zdrojového k=odu

3 TEORETICKÁ VÝCHODISKA

Celá práce je postavena na Esri platformě, čímž jsou chápány produkty společnosti Esri. Esri je americká společnost zabývájící se vývojem software zaměřeného na geografické inforamační systémy. Manželé Dangermondovi ji založili v roce 1969. (http://www.esri.com/about-esri/history)

Z hlediska chápání Esri má GIS tři roviny. První je to GIS jako prostorová databáze reprezentující geografické informace, dále sada map zobrazující prvky a vztahy mezi prvky na zemském povrchu a zároveň i software pro GIS jako sada nástrojů pro odvozování nových informací ze stávajících. Esri tyto tři pohledy na GIS propojuje v softwaru ArcGIS jakožto kompletního GIS, který se skládá s katalogu (kolekce geografický datových sad), map a sad nástrojů pro geografické analýzy.

Esri vytváří integrovanou sadu softwarových produktu ArcGIS, který poskytuje nástroje na kompletní správu GIS a přizpůsobuje produkty různým úrovním nasazení. Výběr produktu záleží na tom, zda zákazník požaduje jedno nebo více uživatelský systém, zda se má jednat o stolní systém nebo server, popř. zda má být dostupný prostřednictvím internetu, nabízí také produkty vhodné pro práci v terénu. (Esri, 2006)

Základní produkty jsou ArcGIS for Desktop ve verzích Basic, Standard, Advanced (arcgis for me http://www.esri.com/software/arcgis/about/gis-for-me), dále ArcGIS for Server (pro Linux a Windows) ve třech úrovních fukncionality (Basic, Standard, Advanced) a dvou úrovních kapacity serveru (Workgroup a Enterpise), ArcGIS for Mobile ve verzích ArcGIS for Windows Mobile, ArcGIS for smartphone a ArcPad, a ArcGIS Online. K tomu všemu Esri přidává velké množství extenzí a další verzí. Kompletní seznam na oficiálních webových stránkách: http://www.esri.com/products nebo http://www.arcdata.cz/produkty-a-sluzby/software/arcgis/.



ArcGIS for Desktop	Basic	Standard	Advanced
ArcGIS for Server	Basic	Standard	Advanced
ArcGIS for Mobile	ArcGIS for Windows Mobile	ArcPAD	ArcGIS for smartphone
ArcGIS Online			

Obr. 1: Verze programu ArcGIS

Nativním formátem programu ArcGIS je geodatabáze. Jedná se o datové uložitě, které umožňuje uložení, přístup a správu dat. Společnost ESRI rozlišuje tři druhy geodatabáze. První dva typy podporují pouze jednoho editujícího uživatele a mnoho uživetelů s právem čtení. Nepodporují dlouhé transakce ani verzování. Tato práce se zaměřuje na třetí typ, kterým je geodatabáze ArcSDE, která se používá pro víceuživatelský přístup, umožňuje dlouho verzování, replikace i verzování (Esri, 2006). Více informací v kapitole 3.1.3 ArcSDE geodatabase.

Typy geodatabází v Esri:

- Personální geodatabáze ukládá data do datového souboru Microsoft Access (.mdb).
- Souborová geodatabáze ukládá data do složek v souborovém systému (.gdb). Je rychlejší a efektivnější než personální geodatabáze.
- Geodatabáze ArcSDE je navržena pro velké projekty s více uživateli. Data se ukládají do vybraného relačního databázového systému.

ArcSDE	Personální	Souborová	ArcSDE
Uložiště	Microsoft Access (.mdb)	Složky v souborovém systému (.gdb)	Relační databázový systém (př. Oracle, MS SQL Server, PostgreSQL)
Licence	ArcGIS for Desktop (Basic, Standard, Advanced)	ArcGIS for Desktop (Basic, Standard, Advanced)	ArcGIS for Desktop (Standard, Advance), ArcGIS for Server
Počet uživatelů	1	1	1 - neomezeně
Dlouhé transakce, verzování, replikace	ne	ne	ano

Obr. 2: Přehled typů geodatabází používaných programem ArcGIS, možno doplnit z http://resources.arcgis.com/en/help/main/10.1/index.html#//003n00000007000000

3.1 Použité programové prostředky

Dnešní pokročilé relační databázové systémy umí víc než jen uložení dat, zajišťují také správu dat, omezují, jaký typ dat je možno vložit do databáze, a zjednodušují získávání dat ze systému (Vieira, 2009).

3.1.1 SQL Server Express 2008

Microsoft SQL Server je relační databázový systém vyvíjený společností Microsoft a dostupný pro různé verze operačního systému Windows. Dodává se v mnoha verzích, které lze nainstalovat na různé hadrwarové platformy na základě odlišných licenčních modelů. SQL Server nabízí 6 základních verzí: Mobile, Express, Workgroup, Standard, Enterprise a Developer Edition. (Whalen, 2008)

Enterprise edition podporuje naprosto vše, co SQL Server nabízí.

Prostorová data jsou implemontována jako CLR rozšíření, přidávají dva prostorové datové - geometry a geography. První jmenovaný slouží k reprezentaci dat, jako např. bodů, čar (linií), polygonů v rovině s omezením souřadnic (konečný prostor). Naproti tomu datový typ geography slouží k reprezentaci stejných dat, avšak na povrchu zeměkoule. Oba typy pracují ve dvou dimenzích (např. na zeměkouli se nebere v potaz výška). Podporuje také indexování dat, index je tvořen standardním B stromem. (Činčura, 2009)

SQL server je podporován a používán Esri platformou od začátku jejího vývoje.

Replikace v MSSQL - co umí, jaké má nástroje, je zdarma?

Struktura pro každou databázi

- co to je
- podpora SDE: SQL server od začátku
- umí databáze replikaci a synchronizaci?
- musí se doinstalovat něco, aby umělo RS nebo je to v defaultu?
- funguje RS jen změnou konfiguráku po instalaci, nebo je potřeba ovlivnit už před instalací?
- jak pracují s geodaty?
- databáze podporuje Esri geodata,
- MS SQL umí datové typy, 2008 express je zadarmo, ale s omezením, nebo koupit databázi SQL + SDE,

které se prodává jen k serverové verzi.

- Postgre používá souřadnicové systémy, SQL omezeně
- OGR?
- datové typy

3.1.2 PostgreSQL 9.x (PostGIS)

PostgreSQL je objektově-relační databázový systém s otevřeným zdrojovým kódem dostupný na většině platforem. Je volně k dispozici pro použití, modifikaci a znovu rozšíření způsobem, který si sami zvolíme. Jedná se o robustní, výkonný, bezpečný, kompatibilní, interoperabilní software s podporou a dobře komentovaným zdrojovým kódem, vyhovuje standardům SQL od verze SQL 2008. PostgreSQL je založen na architektuře klient-server a nabízí velké množství pokročilých funkcí.

S vývojem PostgreSQL začala University of California v Berkley již více než 20 let a není je vyvíjeno a udržováno velkou komunitou nezávislých vývojářů. Používá licenci TPL (The PostgreSQL Licence), která je mírně odlišná od open-source licence BSD (Berkeley Distribution Software), ze které vychází (Riggs, Krosing, 2010).

Řadí se mezi nejpokročilejší databáze díky schopnosti pracovat s velkými objemy dat a díky své rychlosti a funkcionalitě může soupeřit i s populárními komerčními systémy jako je Oracle, IBM DB2, Microsoft SQL Server 2008 a dalšími. (PostgreSQL, 2012).

Samotné PostgreSQL neobsahuje datové typy a funkce pro správu prostorových dat. K tomu je nutné přidat nástavbu PostGIS, která rozšiřuje databázi PostgreSQL o podporu geografických dat. PostGIS implementuje specifikaci "Simple Features for SQL" konsorcia OGC. PostGIS umožňuje ukládání geometrických objektů (bod, linie, polygon), prostorové funkce pro určení vzdáleností, délky linií, výměr a obvodu ploch, výběr indexu při spojení prostorových a atributových dotazů a další.

PostGIS používá dva základní prostorové datové typy geography a geometry. Typ geography ukládá souřadnice v kartézských rovinných souřadnicích, kterým odpovídá souřadnicový systém WGS84. Při výpočtu vzdálenosti dvou bodů tento datový typ vrátí jako výsledek nejkratší vzdálenost v rovině v kilometrech. Typ geometry ukládá data v polárním rovinném systému a umožňuje nastavit souřadnicový systém podle potřeb. Výsledkem dotazu na vzdálenost dvou bodů bude úhel ve stupních. Po převodu do metrické soustavy tedy dostaneme nejkratší vzdálenost na kouli. Rozhodující při výběru datového typu může být například počet funkcí, kterých typ geometry poskytuje mnohem více než geography nebo na velikosti daného území (OpenGeo, 2012).

Existuje také další nástavba, která rozšiřuje ukládání a manipulaci s rastrovými daty (PostGIS Raster), nástavba pro topologickou správu vektorových dat (PostGIS Topology) a pro síťové analýzy (pgRouting). PostGIS je podporován velkou řadou softwarů zabývajících se sravou geografických dat, což také umožňuje snadnou přenositelnost a použitelnost nástavby (příklad software s PostGIS: QGIS, GvSIG, GRASS).

PostGIS používá mnoho knihoven jako GEOS (Geometry Engine Open Source) pro implementaci jednoduchých prostorových prvků a metod pro topologii, PROJ4 pro převod mezi kartografickými projekcemi, GDAL/OGR (Geospatial Data Abstraction Library) pro převod mezi různými vektorovými i rastrovými formáty (Obe, Hsu, 2011). PostGIS má přes 800 funkcí, typů a prostorových indexů (Obe, Hsu, 2012).

PostgreSQL podporuje replikaci i synchronizaci bez nutnosti další instalace. - funguje RS jen změnou konfiguráku po instalaci, nebo je potřeba ovlivnit už před instalací? Replikace v PostgreSQL - co umí, jaké má nástroje, je zdarma, umí vše

PostgreSQL je oficiálně podporovaná databáze pro ukládání geodat v prostředí Esri platformy od verze ArcGIS 9.3. Při instalaci je potřeba zajistit verze, které budou vzájemně kompatibilní. Pro verzi ArcGIS 10.1 jsou podporované verze PostgreSQL 9.0 a PostGIS 1.5., pro ArcGIS 10.1 SP1 (Service Pack 1) je to PostgreSQL 9.1.3 a PostGIS 2.0 (OSGEO, 2013). Kompletní přehled na webu http://trac.osgeo.org/postgis/wiki/UsersWikiPostgisarcgis.

(http://resources.arcgis.com/en/help/system-requirements/10.1/index.html#//015100000075000000)

ArcSDE podporuje pouze datový typ PostGIS Geometry a přidává vlastní datový typ Esri St_Geometry. Výhodou použivání Esri St_Geometry je nezávislost na databázovém systému, tedy snazší přenostitelnost celého řešení. Databáze PostgreSQL se dá použít dvojím způsobem, buď jen jako uložiště dat bez přidání geografického datového typu, nebo včetně datového typu, tedy včetně PostGIS knihovny.

Byly použity verze PostgreSQL 9.1.4 a PostGIS 2.0. Podrobnosti na oficiálních stránách http://www.postgresgl.org/ a http://postgis.refractions.net/.

3.1.3 ArcSDE geodatabase

ArcSDE je technologie firmy Esri pro správu geoprostorových dat uložených v relačních databázových systémem. Jedná se o otevřenou a interoperabilní technologii, která podporuje čtení a zápis mnoha standardů. Využívá jako své nativní datové struktury standard konsorcia OGC: binární jednoduché prvky (Simple Feature) a prostorový typ ISO pro databázové systémy Oracle, IBM DB2 a Informix. Poskytuje vysoký výkon a je přizpůsobena velkému počtu uživatelů (Esri, 2006).

ArcSDE je prostředník pro komunikaci mezi klientem (př. ArcView) a SQL databází (př. PostgreSQL). Umožňuje přístup a správu dat v databázi, současnou editaci jedné databáze více uživateli, zajišťuje prostorový datový typ (St. Geometry), dále integritu dat, dlouhé transakce a práci s verzemi. (Law, 2008).

Existují tři úrovně ArcSDE databáze: desktop (ArcSDE Desktop), skupinová (ArcSDE Workgroup) a podniková (ArcSDE Enterprise). Každé verze má jiné parametry a umožňuje různou úroveň editace, viz obr. 1 (níže).

Technologie ArcSDE vyžaduje dvě úrovně: databázovou a aplikační, která se skládá z ArcObjects a ArcSDE. Databázová úroveň zajišťuje jednoduchý, formální model pro uložení a správu dat ve formě tabulek, definici typů atributů (datových typů), zpracování dotazů či víceuživatelské transakce (Law, 2008). ArcSDE podporuje databázové systémy IBM DB2, IMB Informix, Oracle, Microsoft SQL, PostgreSQL (Esri, 2013a).

Od verze ArcGIS 9.2 je ArcSDE Desktop spolu s databázovým systémem MS SQL Server Express součástí licence produktů ArcGIS for Desktop Standard a Advanced. Takovou databázi mohou současně používat 4 uživatelé, z toho jen jeden může databázi editovat, jsou však omezeni velikostí databáze.

Součastí licence ArcGIS for Server Workgroup je ArcSDE Workgroup, který se liší od verze Desktop především tím, že počet uživatelů, kteří mohou součastně editovat nebo prohlížet databázi, se zvýšuje na deset.

Nejvyšší úroveň, ArcSDE Enterprise, je možno získat s licencí ArcGIS for Server Enterprise, která uživatelům přináší nejméně omezení. Mohou si vybrat z několika komerčních i nekomerčních databázových systémů, počet uživatelů není omezen, stejně jako velikost databáze.

K ArcSDE a vybrané databázi je možno přistupovat přes ArcCatalog, není tedy potřeba instalace dalšího software nebo zkušenost s administrací databáze (Esri, 2006).

Replikaci a synchronizaci dat umožňují pouze ArcSDE Enterprise a Workgroup (Esri, 2013b).

	Enterprise	Workgroup	Desktop
Application Scenario	Large-scale enterprise application scenarios	Small- to medium-sized departmental application scenarios	Small teams or a single user who requires the <u>functionality of a</u> <u>multiuser geodatabase</u>
Data Storage	Enterprise RDBMS Platform DB2 Informix Oracle PostgreSQL SQL Server	SQL Server Express	SQL Server Express
Management Interface	ArcCatalog RDBMS ArcSDE command line	ArcCatalog	ArcCatalog
Storage Capacity	Depends on the server	10 GB	10 GB
Licensing Availability	ArcGIS for Server Enterprise	ArcGIS for Server Workgroup	ArcGIS Engine ArcGIS for Desktop Advanced ArcGIS for Desktop Standard
Supported OS Platform	Any platform	Windows	Windows
Number of Concurrent Users	Unlimited editors and readers	10 editors and readers	1 editor and 3 readers

Obr. 3. Přehled verzi ArcSDE, jejich parametrů a možností

Správa transakční geodatabáze - Esri 2006 str. 109

3.2 Použité postupy

Tato kapitola detailně rozebírá procesy synchronizace a replikace, včetně popisu prostředků, které pro synchronizaci a replikaci využívají ESRI produkty.

Popis požadavků, možností a předpokladů pro správný průběh těchto procesů.

3.2.1 Replikace

Replikace je proces, u kterého jsou data distribuována na více než jeden databázový server (Whalen, 2008). Existují dva nejčastější důvody pro replikaci. Prvním je dostupnost dat neboli snížení pravděpodobnosti,

že data budou nedostupná. To zajistíme tím, že bude úplná kopie dat na více než jednom systému. Druhým je pohyb dat. (Data movement: Allowing data to be used by additional applications or workloads on additional hardware. Examples are Reference Data Management, where a single central server might provide information to many other applications, and also Business Intelligence/Reporting Systems.) (Riggs, Krosing, 2010).

Z pohledu společnosti Esri je replikace geodatabáze chápaná jako rozdělení dat do dvou a více databází způsobem, který umožňuje synchronizovat změny, které v datech nastanou. Jedná se o asynchronní replikaci, při které replikované databáze funguje samostatně, ale stále se synchronizují změny. Vzhledem k tomu, že proces replikace je implementován pří do ArcObjects a ArcSDE, nezáleží na konkrétním databázovém systému (Law, 2008).

MS SQL Server rozlišuje několik typů replikace.

http://technet.microsoft.com/cs-cz/sqlserver/cc510303.aspx

- Prvním je snímková replikace, která dovoluje pořídit snímek dat a načíst jej do další databáze. Tato
 funkce je mimořádně užitečná, máte-li relativně malou tabulku s daty, která se mění zřídka, z chcete ji
 přesunout kvůli tvorbě sestav.
- Dalším typem je transakční replikace, která umožňuje provádět přírůstkové změny na sekundárním serveru. Změny primární tabulky (tzv. publikující) se aplikují na sekundární tabulku (tvz. odebírající).
- Při slučované replikaci je možné mít samostatné publikující databáze a synchronizovat data mezi nimi. To je výhodně pokud data vznikají na více místech, ale pořád je potřeba mít aktuální databázi. (Whalen, 2008)

pozn: využití - záloha serveru při výpadku, proti ztrátě dat, snížení zátěže serveru, sdílení dvou a více databází

synchroná x asynchronná replikace jednosměrná x obousměrná replikace

Replication types typy replikací v esri

http://resources.arcgis.com/en/help/main/10.1/index.html#//003n000000t60000000

Replice PostgreSQL

http://www.postgresql.org/docs/9.1/static/runtime-config-replication.html

Replicas and geodatabases

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Replicas_and_geodatabases/002700000023 000000/

3.2.2 Synchronizace

3.2.3 Verzování

dle Esri lišta verzování

3.3 Geodata

Co to je, jak se ukládají, kdo je používá a jak (1 odstavec = rešerše klasická) jaké typy, jak vypadá formát pro uložení (Point(X,Y))
norma simple feature
souřadnicové systémy: PostgreSQL používá, SQL omezeně - knihovna OGR
Esri ctí vektor, raster, jeho uložení v databázi (feature dataset - ukázka)

Geodata dle Esri

http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_geodata/019r00000002000000/

4 VÝSLEDKY

5 DISKUZE

6 ZÁVĚR

Zkratky

Esri Environmental Systems Research Institute SDE Spatial Database Engine Esri SQL

ZDROJE

ČINČURA, Jiří. MS SQL 2008 – prostorová data poprvé. *Databázový svět* [online]. 2009 [cit. 2013-08-12]. ESRI. *ArcGIS 9: Co je ArcGIS 9.2?*. United States: ESRI Press, US, 2006. ISBN 15-894-8166-6. (Esri, 2013a) ESRI. A quick tour of working with databases in ArcGIS. *ArcGIS Help 10.1* [online]. 2013 [cit. 2013-08-02]. Dostupné z:

http://resources.arcgis.com/en/help/main/10.1/index.html#/A_quick_tour_of_working_with_databases_in_ArcGIS/019v000000_080000000/

(Esri, 2013b) ESRI. Preparing data for replication. *ArcGIS Help 10.1* [online]. 2013 [cit. 2013-08-02]. Dostupné z: http://resources.arcgis.com/en/help/main/10.1/index.html#/Preparing_data_for_replication/003n000000z5000000/OBE, Regina a Leo HSU. *PostGIS in action*. London: Pearson Education [distributor], 2011, 492 s. ISBN 19-351-8226-9. OBE, Regina a Leo HSU. *PostgresqI: Up and Running*. Sebastopol, CA: O'Reilly, 2012, 164 s. ISBN 978-144-9326-333. (OpenGeo, 2012) OPENGEO. Introduction to PostGIS [online]. 2012b [cit. 2012-08-08]. Section 17: Geography.

Dostupné z: http://workshops.opengeo.org/stack-intro/openlayers.html

(OSGEO, 2013) OSGEO. PostGIS and ArcSDE/ArcGIS Articles. *PostGIS Tracker and Wiki* [online]. 2013 [cit. 2013-08-08]. Dostupné z: http://trac.osgeo.org/postgis/wiki/UsersWikiPostgisarcgis

LAW, Derek. *Enterprise Geodatabase 101: A review of design and key features for GIS managers and database administrators*. Esri: Understanding our world. [online]. 2008, [cit. 2013-06-18]. Dostupné z: http://www.esri.com/news/arcuser/0408/entergdb_101.html

RIGGS, Simon a Hannu KROSING. PostgreSQL 9 administration cookbook: solve real-world PostgreSQL problems with over 100 simple, yet incredibly effective recipes. Birmingham: Packt Publishing, 2010, 345 s. ISBN 978-1-849510-28-8. POSTGRESQL. FAQ - PostgreSQL wiki [online]. 2012 [cit. 2012-08-08]. Dostupné z: http://wiki.postgresql.org/wiki/FAQ VIEIRA, Robert. Professional Microsoft SQL server 2008 programming. Indianapolis, IN: Wiley Pub., 2009, 893 s. Wrox

professional guides. ISBN 04-702-5702-4.

WHALEN, Edward a kol. *Microsoft SQL Server 2005: velký průvodce administrátora*. Vyd. 1. Překlad Jakub Mikulaštík, David Krásenský. Brno: Computer Press, 2008, 1080 s. Administrace (Computer Press). ISBN 978-80-251-1949-5.