

# Synchronizace a replikace geodat v prostředí Esri platformy

## Úvod

## 1 CÍLE PRÁCE

## 2 POUŽITÉ METODY A POSTUPY PRÁCE

## 3 TEORETICKÁ VÝCHODISKA

### 3.1 Vymezení pojmů

### 3.2 Replikace

#### 3.2.2 Synchronizace

#### 3.2.3 Verzování

### 3.3 ArcGIS produkty

### 3.4 Použité programové prostředky

#### 3.3.1 PostgreSQL 9.x (PostGIS)

#### 3.4.2 Microsoft SQL Server Express 2008

#### 3.4.3 ArcSDE geodatabase

### 3.5 Příprava prostředí pro testování

#### 3.4.1 Konfigurace

## 4 NÁVRCH A IMPLEMENTACE REPLIKACE

### 4.1 Návrh možných replikačních řešení

### 4.2 Připojení produktů ArcGIS k databázi PostgreSQL

### 4.3 Implementace replikace

#### 4.3.1 Slony-I replikace

#### 4.3.2. Built-in replikace

#### 4.3.3 pg-pool - podpora load-balancing

#### 4.3.4 SDE

## 5 TESTOVÁNÍ

## 6 DISKUZE

## 7 ZÁVĚR

## Zkratky

## ZDROJE

# Úvod

Dnešní trend je stále více dat ukládat a ponechávat pouze v digitální podobě. Mnoho dokumentů už se vůbec netiskne do papírové podoby. Tento trend dnes podporují i elektronické podpisy, díky kterým je tištěná verze naprosto zbytečná. S přibývajícím počtem dat je však třeba řešit komplikace, které počítačová data přináší. Počítačové experti řeší například kam ukládat tak velké množství dat, jak data aktualizovat, jak zabránit poškození dat ať už způsobených lidským faktorem či poškozením hardwaru. V případě, že se poškodí disk, můžeme často během okamžiku přijít o všechna data, někdy však pro ztrátu dat stačí stiknoutou pouhé jedno tlačítko na klávesnici. Určitě už se vám nejednou stalo, že jste se nemohli přihlásit do svého účtu na internetu z důvodu přetížení serveru. Jak zabránit těmto komplikacím, které mohou poškodit či zcela zničit celou dosavadní práci nebo zrychlit celý proces práce s tady? Řešením velkého počtu výše uvedených problémů je replikace dat. Jedná se pokročilou funkcí, kterou nabízí dnešní databázové servery, zajišťující robustnost databáze a vysokou dostupnost dat tím, že data zkopíruje na více serverů.

Replikaci lze využít ve všech odvětvích, které pracují s daty. Výjimkou tedy není ani geoinformatika, která pracuje s velkým objemem dat, které navíc nesou informaci o geografické poloze. Z mého pohledu data středně velkého nebo velkého projektu je nejvhodnější ukládat do databáze. Nabízí nám to sofistikované uložení dat, snadné propojení jednotlivých vrstev, snadnou přenositelnost dat a další výhody. Replikace se dá využít pro kopii dat a následnou aktualizaci změn. Výhodou databáze je, že se při změně jednoho prvku, aktualizuje v databázi pouze jeden řádek a nekopíruje se znovu celá databáze, což je jednoznačná výhoda oproti binárnímu uložení dat například ve formátu shapefile.

Replikaci ocení i uživatelé, kteří pracují na jednom projektu. Z hlediska rychlosti práce s databází je výhodnější mít databázi přímo na počítači, na kterém pracují, než data in-real time stahovat ze serveru. Po dokončení editace se data replikují prostřednictvím počítačové sítě nebo internetu. Dobrým příkladem využitelnosti replikace je také nový trend využívání offline mobilních aplikací v mobilních telefonech. Databáze se vždy replikuje do mobilního telefonu, vždy když se klient připojí na internetovou síť, aplikace kontroluje zda není na serveru novější verze databáze a pokud je, zkopíruje pouze změny, které proběhly od posledního stahování. (Jako příklad z geoinformatického prostředí bych uvedla diplomovou práci Dalibora Janáka, který řeší replikaci databáze lezeckých cest do mobilní aplikace.) Databázové systémy nabízí širokou škálu nastavitelnosti, která umožňuje přizpůsobit replikaci danému řešení.

# 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 možností dalšího podporovaného systému PostgreSQL.

Můj jeden odstavec - 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.

### POZNÁMKY

- Jakými kritérii se hodnotí, zda replikace probíhá správně - it-literatura.
- co budu sledovat - 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
- 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 sítově
  4. přistupuje se online
- zdroje: OGC, knihy, BostonGIS.com, dbsvět.cz (Co umí server 2008 - česky), novinky 2012 (opravy, co je nového)
- 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, datové sady, geodatabáze, modely, metadata a služby (Esri, 2006)."

### 3 TEORETICKÁ VÝCHODISKA

Jak definuje Oppel (2009), databáze je soubor vzájemně propojených datových položek, které jsou spravovány jako jeden celek (Oppel, 2009). Databáze představuje entity, atributy a logické vztahy mezi entitami, často zvané relace. Jinými slovy, databáze obsahuje data, která logicky souvisejí (Connolly, 2005). Databáze umožňuje ukládání a editaci dat, rychlé vyhledávání dat a komplexní analýzu dat (Mimjian, 2001). Systém řízení báze dat<sup>1</sup> je počítačový software, který umožňuje uživatelům přistupovat k databázi, definovat, vytvářet a udržovat data (Connolly, 2005). Pro uložení dat malého projektu je samozřejmě možno použít tabulkového procesoru či jiného formátu určeného pro ukládání dat. Pro komplexní správu dat velkého projektu je však databáze více než vhodná.

Prostorová databáze, někdy také zvaná geodatabáze, není nic jiného než databáze připaná o datový typ určený pro ukládání prostorové informace o prvku, prostorové indexy a sadu funkcí vhodných pro správu prostorových dat. Více informací o prostorových databázích v kapitole 3.3.1 PostgreSQL 9.x (PostGIS) a 3.3.2 MS SQL Server 2008.

Z toho vyvstává otázka, co jsou prostorová data, také zvaná geodata. Z pohledu společnosti ESRI se jedná se prvky, které nesou informaci o geografické poloze, zakódovanou informaci o tvaru (bod, line, polygon) a popis geografického jevu. Tato geodata jsou uložena ve formátu, který je možno použít v geografickém informačním systému (ArcGIS, 2009). Příkladem takového formátu může být vektorový Esri shapefile, Esri coverage, GML, KML nebo rastrový Erdas Image a GeoTIFF. Dalším způsobem je již zmíněná databáze, do níž se vektorová data ukládají ve specifickém tvaru daném standardem OGC<sup>2</sup> Simply Feature for SQL 1.2.1, který specifikuje způsob uložení dat v digitální podobě. Simple Features je založen na 2D geometrii s možností lineární interpolace mezi lomovými body. To umožňuje vložení následujících prvků:

- POINT(0 0)
- LINESTRING(0 0, 1 1, 1 2)
- POLYGON ((0 0, 4 0, 4 0, 4 0, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1))
- MULTIPOINT((0 0), (1 2))
- MULTILINESTRING((0 0, 1 1, 1 2), (2 3, 3 2, 5 4))
- MULTIPOLYGON(((0 0, 4 0, 4 0, 4 0, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1)), ((-1 -1, -1 -2, -2 -2, -2 -1, -1 -1)))
- GEOMETRYCOLLECTION(POINT(2 3), LINESTRING(2 3, 3 4))<sup>3</sup>

Hodnoty lze dále vkládat přes Well-Known Binary (WKB) nebo Well-Known Text (WKT) reprezentaci:

- bytea WKB = ST\_AsBinary(geometry);
- text WKT = ST\_AsText(geometry);

Ukázka formátu WKT pro ukládání dat do PostGIS databáze:

```
(LINESTRING(15.96876 50.85494, 17.21879 49.60494, 18.27348 49.85350), 5514))
```

 (1)

První slovo určuje druh prvku (point, linestring, polygon, multibod, multipoint, ...), následují v závorce vypsane souřadnice lomových bodů. Posledním volitelným parametrem je kód souřadnicového systému.

Esri ctí vektor, raster, jeho uložení v databázi (feature dataset - ukázka) udělat printscreen

#### 3.1 Vymezení pojmů

Pro lepší porozumění textu této práce je potřeba definovat pojmy replikace, synchronizace a verzování,

<sup>1</sup> v anglickém originále database management system (DBMS)

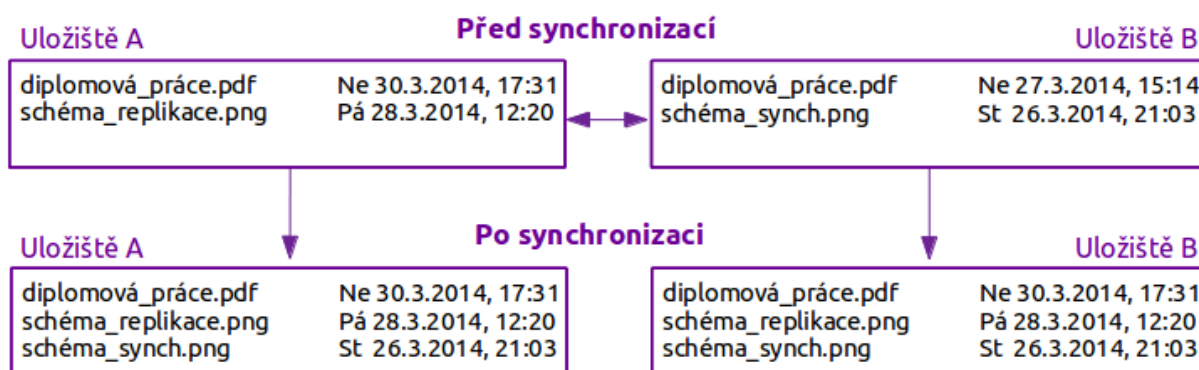
<sup>2</sup> OGC standardy jsou kontrolovány konsorciem Open Geospatial Consortium, zdroj <http://www.opengeospatial.org/ogc>

<sup>3</sup> zdroj [http://postgis.net/docs/manual-2.1/using\\_postgis\\_dbmanagement.html#RefObject](http://postgis.net/docs/manual-2.1/using_postgis_dbmanagement.html#RefObject)

včetně popisu toho, jak jsou dané pojmy chápány v produktech ArcGIS. Je vhodné upozornit, že výše zmíněné procesy jsou v literatuře často chápány lehce odlišně. Některé zdroje pojmy replikace a synchronizace rozlišují, jiné je naopak považují za synonyma.

Všechny dotyčné pojmy úzce souvisí se zálohováním dat, tedy kopírováním dat mezi dvěma a více uložšti. To, co tyto pojmy spojuje, je totiž vždy, v nějaké míře, zabránění ztrátě dat, ať už chybou či fyzickým poškozením disku. Dané pojmy se poté liší například konkrétní způsobem provedení zálohy, či důvodem kopírování dat.

Z mého pohledu je synchronizace nadmnožinou replikace. V případě, že existují dva datové zdroje a v jeden okamžik se rozhodneme, že chceme tyto dvě složky sjednotit, poté je možno mluvit o synchronizaci souborů či datových složek. Soubor, který se podle názvu nachází ve složce A a zároveň se nenachází ve složce B, se jednoduše zkopíruje do složky A. U souborů se stejným názvem, se dále porovnává čas posledního zápisu, velikost nebo obsah souboru. Poté je soubor se starším datem, resp. menší velikostí, přepsán tím novějším, resp. větším. Synchronizací se tedy dá proces označit v okamžiku, kdy existují nejméně dva datové zdroje a smyslem synchronizace je porovnat dvě a více uložště a dostat je do stejného stavu. To může například přispět snazší spolupráci více uživatelů nad stejnými daty.



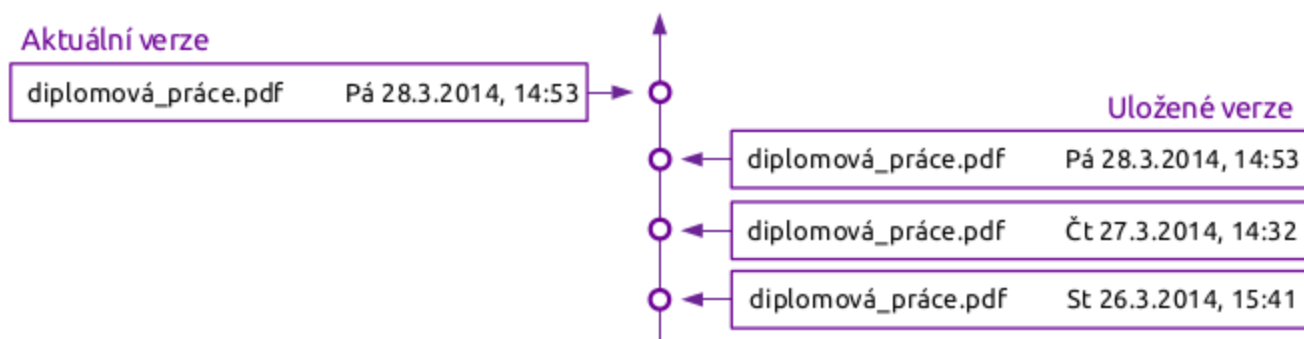
Obr. č: Příklad obousměrné synchronizace dat mezi dvěma datovými uložšti

Replikace naopak, podle mého názoru, začíná s daty existujícími pouze na jednom uložšti. Často je tento proces používán právě ve spojitosti s databázemi, kdy je kopie dat (také replika) tvořena z důvodu snížení zátěže serveru, či ochrany dat. V případě že je tato kopie již vytvořena, je poté možno mluvit i o synchronizaci dat, protože replika průběžně kontroluje, zda na hlavním serveru nedošlo ke změně, a pokud ano, dané změny zkopíruje.

Oba procesy je možno použít jednostranně, tedy kopírovat data pouze z jednoho uložště na druhé a nikoliv opačně, nebo oboustranně, kdy se data kopírují navzájem mezi sebou.

Specifickým způsobem zálohy dat je verzování, kdy se data na záložním datovém uložšti nepřepisují, ale systematicky ukládají v takzvaných verzích tak, aby se uživatel mohl snadno kdykoliv vrátit k předchozím stavům souborů. Smyslem verzování je zachovat všechny zvolené stavy práce, čímž se verzování liší od zálohování, kde stačí mít aktuální kopii daných dat. Verzování může probíhat ručně, poloautomatizovaně či plně automatizovaně díky speciálním nástrojům pro strávu verzí. Oblíbeným verzovací systémem programátorů je GIT<sup>4</sup>, open-source nástroj pro správu verzí, který pomáhá při práci s malými i velkými projekty a podporuje týmovou spolupráci. Umožňuje vrátit jednotlivé soubory nebo celý projekt do předchozího stavu, porovnávat změny provedené v průběhu času, zjistit, kdo naposledy upravil něco, co nyní možná způsobuje problémy, kdo vložil jakou verzi a mnoho dalšího (Chacon, 2009). GIT je velice vhodný pro textové soubory, protože dokáže analyzovat části textu, či programového kódu a zvýraznit, která část se změnila.

<sup>4</sup> více na <http://git-scm.com/>

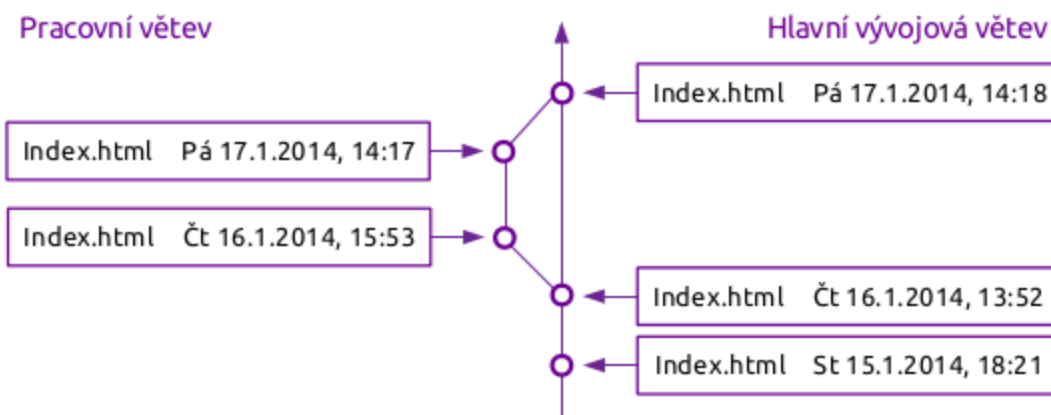


Obr. č: Příklad verzování souboru

Samotná databáze verzování dat neumožňuje. Nejsnazším způsobem, jak získat verze dat, je dump, tedy export databáze do souboru. Takový soubor se poté může verzovat podobným způsobem jako jakýkoliv jiný binární soubor typu shapefile. A to samé platí i pokud v databázi ukládáme geodata.

Proto byl vytvořen verzovací systém také pro prostorová data, který vychází ze systému git a nese název GeoGIT. Umožňuje uživatelům zachovávat změny v souborech shapefiles, SpatialLite a z databáze PostGIS (PostgreSQL). Umožňuje, tak jako git, uchovávat historii prostorových dat, či vrátit se k předchozí verzi.

Verzováním může být chápáno také jako vytvoření pracovní verze v případě, že programový kód či data jsou plně funkční či aktuální, tak aby nedošlo k poškození této správné verze. Pracovní verze je kopie aktuálního souboru, na které je možno pracovat a zkoušet, v případě, že práce nedopadne podle představ, je možno změny zahodit, pokud je tomu naopak, je možno pracovní verzi sjednotit s platnou verzí. Tento způsob verzování umožňuje GIT i GeoGIT.



Obr. č: Příklad verzování souboru s použitím pracovní větve

Až doposud byly procesy popisovány z co nejobecnějšího hlediska. Společnost Esri však některé z pojmů chápe odlišně. Dle Law (2008)

## 3.2 Replikace

### MRKNI DO KNÍŽKY MYSQL HIGH AVAILABILITY

Z pohledu společnosti Esri je replikace geodatabáze chápána jako rozdělení dat do dvou a více databází způsobem, který umožňuje synchronizovat změny, které v datech nastanou (Law, 2008). Replikace je tedy proces, u kterého jsou data a databázové objekty kopírovány z jednoho databázového serveru na druhý a poté synchronizované pro zachování souladu obou databází. Synchronizací v tomto případě myslíme kopírováním všech změn, které v databázi nastanou. Použitím databáze je možno data distribuovat na různě vzdálená místa

nebo mezi mobilní uživatele v rámci počítačové sítě a internetu. (Microsoft, 2013).

Existují dva nejčastější důvody pro replikaci. Prvním je dostupnost dat, resp. snížení pravděpodobnosti, že data nebudou dostupná. To může být způsobeno například výpadkem serveru či ztrátou dat. Druhým je škálovatelnost dat, tedy snížení zátěže serveru tím, že se vytvoří úplná kopie dat na jiný server (Obe, Hsu, 2012). Podle Bella (2010) moderní aplikace mají často více čtenářů než zapisovatelů, proto je zbytečné, aby se všichni čtenáři připojovali na Master databázi a zpomalovali tím práci zapisovatelů. Pokud se vytvoří kopie na více serverů, rozdělí se zátěž mezi více serverů a při velkém počtu připojení nedojde ke zpomalení (Bell, Kindahl, Thalmann, 2010). Mechanismus replikace je často velmi flexibilní a dá se dobře konfigurovat, je tedy možné zvolit vhodný způsob replikace podle typu projektu (Whalen, 2008).

Replikace používá dvě základní role, nejčastěji nazývané Master a Slave. Master server nebo pouze Master je server, který poskytuje data k replikaci a probíhají na něm všechny aktualizace. Může být také nazýván Primary server, Provider, Sender, Parent nebo Source server. Naprosto jiný pojem zavádí MS SQL Server, který tento zdrojový server nazývá Publisher (česky Vydavatel). Druhý databázový server je nejčastěji nazýván Slave, Standby, Reciever, Child nebo Subscriber (česky Odběratel). Poslední pojem je také používán MS SQL Serverem. Na tento server se data a aktualizace z Master serveru kopírují (Riggs, Krossing, 2010). Slave server je dostupný vždy jen pro čtení dat, není možné na něj zapisovat změny.

Dříve než se začne cokoli nastavovat, musí se rozhodnout, zda replikace bude synchronní nebo asynchronní. Synchronní replikace znamená, že na Master serveru nikdy neproběhne nová transakce, dokud se poslední transakce neprovedou na Slave serverech. Tento přístup zajistí, že žádná data nebudou v průběhu transakce ztracena. Avšak tento přístup může velice zpomalit přístup k databázi, protože se bude čekat na každou nedokončenou transakci. Druhým způsobem je asynchronní replikace, při které se nová data zapisují na Master server, přestože ještě nedošlo k replikaci stávajících dat na Slave server (Obe, Hsu, 2012). To je možné díky souboru Binary Log (nebo pouze Binlog), do kterého se zapisují všechny změny, které byly provedeny, například přidání záznamu v tabulce) (Bell, Kindahl, Thalmann, 2010).

Single master means that writes can go to exactly one server, which distributes the data to the slaves inside the setup. Slaves may only receive reads but no writes. Multi-master replication allows writes to all the servers inside the cluster (PostgreSQL Replication, 2013) - viz obr. z knížky, str. 13.

Vzhledem k tomu, že proces replikace je implementován při 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 (tzv. odebírající).
- Při **slučované replikaci** je možné mít samostatné publikující databáze a synchronizovat data mezi nimi. To je výhodné 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#//003n000000t6000000>

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/002700000023000000/](http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Replicas_and_geodatabases/002700000023000000/)

### 3.3 ArcGIS produkty

V názvu práce se objevuje spojení Esri platforma, čímž jsou chápány produkty společnosti Esri. Esri je americká společnost zabývající se vývojem software zaměřeného na geografické informační systémy. Manželé Dangermondovi ji založili v roce 1969<sup>5</sup>.

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 software ArcGIS jakožto kompletní GIS, který se skládá s katalogu (kolekce geografický datových sad), map a sad nástrojů pro geografické analýzy.

Esri vytváří integrovanou sadu softwarových produktů 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 uživatelských systémů, 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<sup>6</sup> jsou stolní systémy ArcGIS for Desktop ve verzích Basic, Standard, Advanced<sup>7</sup>, dále serverové verze ArcGIS for Server (pro Linux a Windows) ve třech úrovních funkcionality (Basic, Standard, Advanced) a dvou úrovních kapacity serveru (Workgroup a Enterprise). Další produkt ArcGIS for Mobile ve verzích ArcGIS for Windows Mobile, ArcGIS for Smartphone and Tablet a ArcPad je určený především pro práci v terénu. Neposlední řadě verze dostupná skrze internet ArcGIS Online. K tomu všemu Esri přidává velké množství extenzí a dalších verzí<sup>8</sup>.

---

<sup>5</sup> Více info <http://www.esri.com/about-esri/history>

<sup>6</sup> Názvy jednotlivých produktů použitých v tomto odstavci jsou platné od verze ArcGIS 10.1. Starší verze ArcGIS používají jiné názvy, jejichž přehled je možný na stránkách firmy ARCDATA Praha  
<http://www.arcdata.cz/produkty-a-sluzby/software/arcgis/prejmenovani-arcgis/>.

<sup>7</sup> Zdroj <http://www.esri.com/software/arcgis/about/gis-for-me>

<sup>8</sup> Kompletní seznam na oficiálních webových stránkách Esri <http://www.esri.com/products> nebo <http://www.arcdata.cz/produkty-a-sluzby/software/arcgis/>.

Produkt	Verze		
ArcGIS for Desktop	Basic	Standard	Advanced
ArcGIS for Server	Basic	Standard	Advanced
ArcGIS for Mobile	ArcGIS for Windows Mobile	ArcPAD	ArcGIS for Smartphone and Tablet
ArcGIS Online			

Obr. č: Verze programu ArcGIS platné od verze 10.1.

Dle autorů publikace ArcGIS 9: Co je ArcGIS 9.2? (Esri, 2006) je nativním formátem programu ArcGIS geodatabáze a rozlišují tři druhy geodatabáze. Ani v jednom případě se ale nejedná o databázi v pravém slova smyslu, tak jako ji chápeme v kapitole 3.1.1 a 3.1.2. V každém případě však tyto způsoby uložení umožňují uložení, přístup a správu dat. U prvních dvou typů, personální a souborové geodatabáze, se data ukládají do jednoho binárního souboru, kde jsou však ukládána ve stejné struktuře jako v plnohodnotném databázovém serveru. Do takového geodatabáze můžeme uložit více než jednu vrstvu, což je výrazný rozdíl oproti formátu shapefile. Výhodou je dále možnost použití relací, sofistikované dotazování a v neposlední řadě i snadná přenositelnost, protože takováto databáze bude vždy jen jeden soubor obsahující několik vrstev. Oproti tomu shapefile, který obsahuje jen jednu vrstvu, je tvořen minimálně 4 soubory. Oba tyto typy podporují pouze jednoho editujícího uživatele a mnoho uživatelů s právem čtení. Nepodporují dlouhé transakce ani verzování.

Tato práce se více zaměřuje na třetí typ, kterým je technologie ArcSDE, kterou v některých materiálech je možno vidět pojmenovanou jako “geodatabáze ArcSDE”. Nejedná se o geodatabázi, ale spíše o zprostředkovatele komunikace mezi programem ArcGIS a databázovým server. Umožňuje víceuživatelský přístup, verzování i replikaci (Esri, 2006). Tato technologie využívá jako datové uložení některý z již existujících databázových serverů, např. níže popsané PostgreSQL nebo MS SQL server. Více informací o této technologii v kapitole 3.1.3 ArcSDE geodatabase.

databáze		souborová .gdb <sup>1</sup>	personální .mdb <sup>1</sup>
datové uložení/ databázový server		lokální souborový systém	MS Access
licence		ArcGIS for Desktop (všechny verze)	ArcGIS for Desktop (všechny verze)
operační systém		Windows (možná i jiné)	Windows
požaduje ArcSDE		ne	ne
vlastní datový typ		ne	ne
víceuživatelská editace		ano, ale s limity	ne
počet editorů		1 pro každý dataset nebo tabulku <sup>2</sup>	1 <sup>2</sup>
počet čtenářů		více než 1 <sup>2</sup>	více než 1 <sup>2</sup>
replikace	typ replikace	souborová replikace	souborová replikace
	master server	ne <sup>1</sup>	ne <sup>1</sup>
	slave server	ano	ano
verzování		ne	ne
velikostní limity		1TB pro každý dataset <sup>2</sup>	2GB <sup>2</sup>

## 3.4 Použité programové prostředky

Dnešní pokročilé relační databázové systémy umí víc než jen uložení dat, zajišťují kompletní správu dat, určují, jaký typ dat je možno vložit do databáze, a zjednodušují získávání dat ze systému (Vieira, 2009). Práce se zabývá pokročilejší správou databáze a to replikací a synchronizací dat.

### 3.3.1 PostgreSQL 9.x (PostGIS)

PostgreSQL je objektově-relační databázový systém s otevřeným zdrojovým kódem dostupný na většině platform. Je volně k dispozici pro použití, modifikaci a znovu rozšíření způsobem, který si sami zvolíme. Jedná se o robustní, výkonný, bezpečný, kompatibilní a interoperabilní software s podporou a dobře komentovaným zdrojovým kódem. Vyhovuje standardům SQL od verze SQL 2008 a nabízí velké množství pokročilých funkcí. PostgreSQL je založen na architektuře klient-server, to znamená, že server pořád běží a čeká na dotazy klienta (Momjian, 2001).

S vývojem databázového serveru PostgreSQL začala University of California v Berkley již více než 20 let. Nyní je vyvíjen a udržován 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, 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 vhodné 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), použití prostorových funkcí pro určení vzdáleností, délky linií, výměr a obvodu ploch, výběr indexu při spojení prostorových a atributových dotazů a mnoho dalších.

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 kilometrech v rovině. Typ geometry data ukládá v polárním rovinném systému a umožňuje nastavit souřadnicový systém podle potřeb. Výsledkem dotazu na vzdálenost dvou bodů tedy bude úhel ve stupních. Po převodu do metrické soustavy dostaneme nejkratší vzdálenost na kouli. Při výběru datového typu může být rozhodující například počet funkcí, kterých typ geometry poskytuje mnohem více než geography, nebo velikosti daného území. Pro menší území je vhodný typ geography, pro větší území naopak geometry (OpenGeo, 2012).

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

PostGIS používá mnoho běžně používaných knihoven jako GEOS (Geometry Engine Open Source) pro implementaci jednoduchých prostorových prvků a metod pro topologii, PROJ4 pro převod mezi kartografickými projekcemi nebo GDAL/OGR (Geospatial Data Abstraction Library) pro převod mezi různými vektorovými i rastrovými formáty (Obe, Hsu, 2011). PostGIS 1.5. obsahovala přes 800 funkcí, typů a prostorových indexů

(Obe, Hsu, 2012). Aktuální verze PostGIS je 2.1.<sup>9</sup>

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

Od verze ArcGIS 9.3. je PostgreSQL je oficiálně podporovanou databází pro ukládání geodat v produktech ArcGIS. Při instalaci je pouze potřeba zajistit kompatibilitu verzí. 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)<sup>10</sup>. Databáze PostgreSQL se dá v ArcGIS produktech použít dvojím způsobem. Buď jen jako uložení dat bez přidání geografického datového typu, nebo včetně datového typu, tedy včetně PostGIS knihovny. ArcSDE podporuje pouze datový typ PostGIS Geometry a přidává vlastní datový typ Esri St\_Geometry. Výhodou používání Esri St\_Geometry je nezávislost na zvoleném databázovém systému, tedy snazší přenositelnost celého řešení.

Práce byla testována na verzích PostgreSQL 9.1.4<sup>11</sup> a PostGIS 2.0.

### 3.4.2 Microsoft SQL Server Express 2008

Microsoft SQL Server (dále MS SQL Server) je relační databázový systém vyvíjený společností Microsoft dostupný pro různé verze operačního systému Windows. Dodává se v mnoha verzích, které lze nainstalovat na různé hardwarové platformy na základě odlišných licenčních modelů (Whalen, 2008). Podle Leitera (2009) SQL Server nabízí 8 základních verzí: Enterprise, Standard, Workgroup, Web, Express, Express Advanced Edition, Developer Edition a Compact Edition. Enterprise edition podporuje naprosto vše, co SQL Server nabízí, naopak verze Express, která je dostupná zdarma, obsahuje omezení některých funkcí a proto je vhodná spíše pro malé nebo začínající projekty (Leiter, 2009).

Prostorová data jsou implementována jako CLR rozšíření a přidávají databázovému serveru dva prostorové datové typy geometry a geography. Rozdíl mezi datovými typy je podobný jako u PostgreSQL. První jmenovaný slouží k reprezentaci dat (bodů, linií, polygonů) v rovině, naproti tomu datový typ geography slouží ukládání stejných dat na povrchu zeměkoule. Oba typy pracují ve dvou dimenzích (nebývá se v potaz výška). Podporuje také indexování dat, index je tvořen standardním B stromem (Činčura, 2009).

SQL Server je podporován a používán ArcGIS produkty od začátku jeho vývoje. Verze ArcGIS Enterprise může být propojena s jakoukoliv uživatelem zvolenou a zakoupenou licencí databázového systému. Verze ArcSDE Desktop a Workgroup používají verzi Express, která je dostupná zdarma a podporuje většinu základních funkcí. Replikaci plně podporuje verze Enterprise, ostatní verze ji podporují pouze s omezenými funkcemi. Avšak již zmiňovaná verze Express, která je podporována ArcSDE Desktop a Workgroup, může být použita pouze jako slave server, tedy odběratelem replikovaných dat, není tedy možné do takovéto databáze připojené do replikačního clusteru zapisovat. Nemůže být tím, kdo poskytuje data k replikaci (Whalen, 2008).

- funguje RS jen změnou konfiguráku po instalaci, nebo je potřeba ovlivnit už před instalací? Replikace v MSSQL - co umí, jaké má nástroje?

- Postgre používá souřadnicové systémy, SQL omezeně

- OGR?

### 3.4.3 ArcSDE geodatabase

ArcSDE je technologie firmy Esri pro správu geoprostorových dat uložených v relačních databázových systémech. Jedná se o otevřenou a interoperabilní technologii, která podporuje čtení a zápis mnoha standardů.

<sup>9</sup> <http://postgis.refractory.net/>

<sup>10</sup> Zdroj a další informace na stránkách PostgreSQL <http://trac.osgeo.org/postgis/wiki/UsersWikiPostgisarcgis> nebo ArcGIS <http://resources.arcgis.com/en/help/system-requirements/10.1/index.html#/015100000075000000>

<sup>11</sup> <http://www.postgresql.org/>

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).

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, IBM Informix, Oracle, Microsoft SQL, PostgreSQL (Esri, 2013a).

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. 3.

databáze		ArcSDE		
		Desktop	Workgroup	Enterprise
datové uložště/ databázový server		SQL Server Express	SQL Server Express	PostgreSQL, Oracle, SQL Server a další
licence		ArcGIS for Desktop (všechny verze)	ArcGIS for Server Workgroup	ArcGIS for Server Enterprise
operační systém		Windows	Windows	všechny operační systémy
požaduje ArcSDE		ano	ano	ano
vlastní datový typ		ne	ne	ano
víceuživatelská editace		ne	ano	ano
počet editorů		1	10	bez limitu
počet čtenářů		3	10	bez limitu
replikace	typ replikace	databázová replikace	databázová replikace	databázová replikace
	master server	ne	ne	ano
	slave server	ano	ano	ano
verzování		ano	ano	ano
závislost na sítích		lokální síť	lokální síť, internet	lokální síť, internet
velikostní limity		10GB	10GB	záleží na velikosti serveru

Obr. č: Přehled verzí ArcSDE, jejich parametrů a možností

zdroj: <http://www.esri.com/software/arcgis/geodatabase/multi-user-geodatabase>

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.

Součástí licence ArcGIS for Server Workgroup je ArcSDE Workgroup, která se liší od verze Desktop především tím, že počet uživatelů, kteří mohou současně editovat nebo prohlížet databázi, je zvýšen 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). Jak už bylo

zmíněno v předchozí kapitole 3.3.2 Microsoft SQL Server Express 2008, SQL Server Express je možný použít v replikačním clusteru pouze jako slave server.

Správa transakční geodatabáze - Esri 2006 str. 109

## 3.5 Příprava prostředí pro testování

### 3.4.1 Konfigurace

- 2.2.1? podkapitoly - kde konfigurák, jak, výseky zdrojového kódu konfigurace čeho?

## 4 NÁVRCH A IMPLEMENTACE REPLIKACE

### 4.1 Návrh možných replikačních řešení

Porovnání built-in x Slony-I replikace

### 4.2 Připojení produktů ArcGIS k databázi PostgreSQL

### 4.3 Implementace replikace

#### 4.3.1 Slony-I replikace

#### 4.3.2. Built-in replikace

#### 4.3.3 pg-pool - podpora load-balancing

#### 4.3.4 SDE

## 5 TESTOVÁNÍ

pg\_bench

## 6 DISKUZE

## 7 ZÁVĚR

Tato práce navrhuje možné použití replikace pomocí databázového serveru PostgreSQL.

# Zkratky

Esri ou  
SDE Spatial Database Engine  
SQL Structured Query Language

## ZDROJE

### Knihy

(Bell, Kindahl, Thalmann, 2010) BELL, Charles, Mats KINDAHL a Lars THALMANN. *MySQL high availability*. 1. ed. Sebastopol, CA: O'Reilly Media, Inc, 2010. ISBN 978-059-6807-306.

(Connolly, 2005) CONNOLLY, Thomas. *Database Systems: A Practical Approach to Design, Implementation, and Management*. 4th ed. Harlow: Addison-Wesley, 2005, 1374 s. ISBN 03-212-1025-5.

(Esri, 2006) ESRI. *ArcGIS 9: Co je ArcGIS 9.2?*. United States: ESRI Press, US, 2006. ISBN 15-894-8166-6.

(Chacon, 2009) CHACON, Scott. *Pro Git*. Praha: CZ.NIC, c2009, 263 s. Edice CZ.NIC. ISBN 978-80-904248-1-4.

(Leiter, 2009) LEITER, Chris. *Beginning Microsoft SQL server 2008 administration*. Indianapolis, IN: Wiley Pub., 2009, p. cm. ISBN 978-047-0440-919.

(Momjian, 2001) MOMJIAN, Bruce. *PostgreSQL: introduction and concepts*. Boston, MA: Addison-Wesley, 2001, xxviii, 461 p. ISBN 02-017-0331-9.

(Obe, Hsu, 2011) OBE, Regina a Leo HSU. *PostGIS in action*. London: Pearson Education [distributor], 2011, 492 s. ISBN 19-351-8226-9.

(Obe, Hsu, 2012) OBE, Regina a Leo HSU. *Postgresql: Up and Running*. Sebastopol, CA: O'Reilly, 2012, 164 s. ISBN 978-144-9326-333.

(Oppel, 2009) OPPEL, Andrew J. *Databases: a beginner's guide*. New York: McGraw-Hill, 2009, xviii, 478 p. ISBN 00-716-0846-X.

(Riggs, Krossing, 2010) 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.

(Vieira, 2009) VIEIRA, Robert. *Professional Microsoft SQL server 2008 programming*. Indianapolis, IN: Wiley Pub., 2009, 893 s. Wrox professional guides. ISBN 04-702-5702-4.

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

### Internetové zdroje

(Činčura, 2009) ČINČURA, Jiří. *MS SQL 2008 – prostorová data poprvé. Databázový svět* [online]. 2009 [cit. 2013-08-12].

(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/019v00000008000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/A_quick_tour_of_working_with_databases_in_ArcGIS/019v00000008000000/)

(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/](http://resources.arcgis.com/en/help/main/10.1/index.html#/Preparing_data_for_replication/003n000000z5000000/)

(Law, 2008) 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](http://www.esri.com/news/arcuser/0408/entergdb_101.html)

(Microsoft, 2013) MICROSOFT. *SQL Server - Replication*. Microsoft [online]. 2013 [cit. 2013-08-27]. Dostupné z: [http://technet.microsoft.com/en-us/library/ms151198\(v=sql.100\).aspx](http://technet.microsoft.com/en-us/library/ms151198(v=sql.100).aspx)

(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>

(PostgreSQL, 2012) POSTGRESQL. *FAQ - PostgreSQL wiki* [online]. 2012 [cit. 2012-08-08]. Dostupné z: <http://wiki.postgresql.org/wiki/FAQ>