

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

Markéta SOLANSKÁ

SYNCHRONIZACE A REPLIKACE GEODAT V PROSTŘEDÍ ESRI PLATFORMY

Magisterská práce

Vedoucí práce: doc. RNDr. Vilém Pechanec, Ph.D.

Olomouc 2014

Čestné prohlášení

Prohlašuji, že jsem magisterskou práci magisterského studia oboru Geoinformatika vypracovala samostatně pod vedením RNDr. Viléma Pechance, Ph.D.

Všechny použité materiály a zdroje jsou citovány s ohledem na vědeckou etiku, autorská práva a zákony na ochranu duševního vlastnictví.

Všechna poskytnutá i vytvořená digitální data nebudu bez souhlasu školy poskytovat.

V Olomouci 4. dubna 2014

Markéta SOLANSKÁ

Ráda bych poděkovala vedoucímu práce doc. RNDr. Vilému Pechancovi, Ph.D. za podněty a připomínky při vypracování práce.

Děkuji také konzultantu Tomáši Vondrovi za pomoc při pochopení a praktickém použití databázového serveru PostgreSQL, za jeho cenné rady a připomínky, stejně tak jako i jeho kolegovi Pavlovi Stěhule.

Dále děkuji konzultantům Boudewijn van Leeuwen a Zalan Tobak působících na Universitě v Szegedu v Maďarsku za inspirativní připomínky a podněty k této práci.

Obsah

ÚVOD	7
1 CÍLE PRÁCE	9
2 POUŽITÉ METODY A POSTUPY PRÁCE	10
2.1 Obrázkygt	10
3 TEORETICKÁ VÝCHODISKA	12
3.1 Vymezení pojmů	13
3.2 Replikace	16
3.3 ArcGIS produkty	19
3.4 Použité programové prostředky	21
3.4.1 PostgreSQL 9.x (PostGIS)	21
3.4.2 Microsoft SQL Server Express 2008	23
3.4.3 ArcSDE geodatabase	24
4 VÝSLEDKY	26
5 DISKUZE	27
6 ZÁVĚR	28
LITERATURA	29
SUMMARY	31
PŘÍLOHY	32

Seznam obrázků

1	Povrch socioekonomického (a) a fyzickogeografického (b) ukazatele . .	11
2	Příklad obousměrné synchronizace dat mezi dvěmi datovými uložišti .	14
3	Příklad obousměrné synchronizace dat mezi dvěmi datovými uložišti .	14
4	Příklad verzování souboru	15
5	Příklad verzování souboru s použitím pracovní větve	16
6	Srovnání Master-Master a Master-Slave replikace	17
7	Rozdíl mezi synchronní a asynchronní replikací	18
8	Ukázka kaskádové replikace	19
9	Verze programu ArcGIS platné od verze 10.1.	20
10	Přehled rozdílů personální a souborové geodatabáze v ArcGIS	21
11	Přehled verzí ArcSDE, jejich parametrů a možností	25

ÚVOD

Úvod je v rozsahu jedné strany. Je důležitou součástí práce. Uvádí do problematiky řešené v bakalářské / magisterské práci. Student v něm vyjadřuje potřebu řešení zadaného tématu, případně návaznost na jeho předcházející práce. Doporučuje se, aby byl do své konečné podoby dopsán až jako poslední, tj. až po napsání celého textu.

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. Počítačovní 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 geoformatika, která pracuje s velkým počtem 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. 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 určitě i uživatelé, které 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

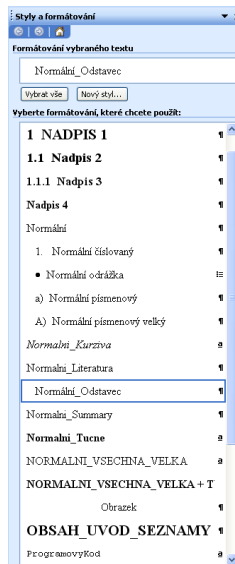
Cílem diplomové práce je provést rešerši a na jejím základě prakticky otestovat proces synchronizace a replikace geodat, které se dnes objevují napříč platformou Esri. V teoretické části práce bude detailně analyzován proces synchronizace a replikace ve všech možných variantách (jednosměrná, dvousměrná, synchronní, asynchronní, ...) a popsány prostředky, které se na platformě Esri k těmto procesům využívají. Rozbor zahrne celé portfolio 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 odstaveček - něco jako - jak vidím vlastní přínos do tématu.

2 POUŽITÉ METODY A POSTUPY PRÁCE

2.1 Obrázky



Esri SQL bla bla bla (viz obr. ??). Pokud chceme uvést překlad z angličtiny můžeme to udělat takto (angl. english words).

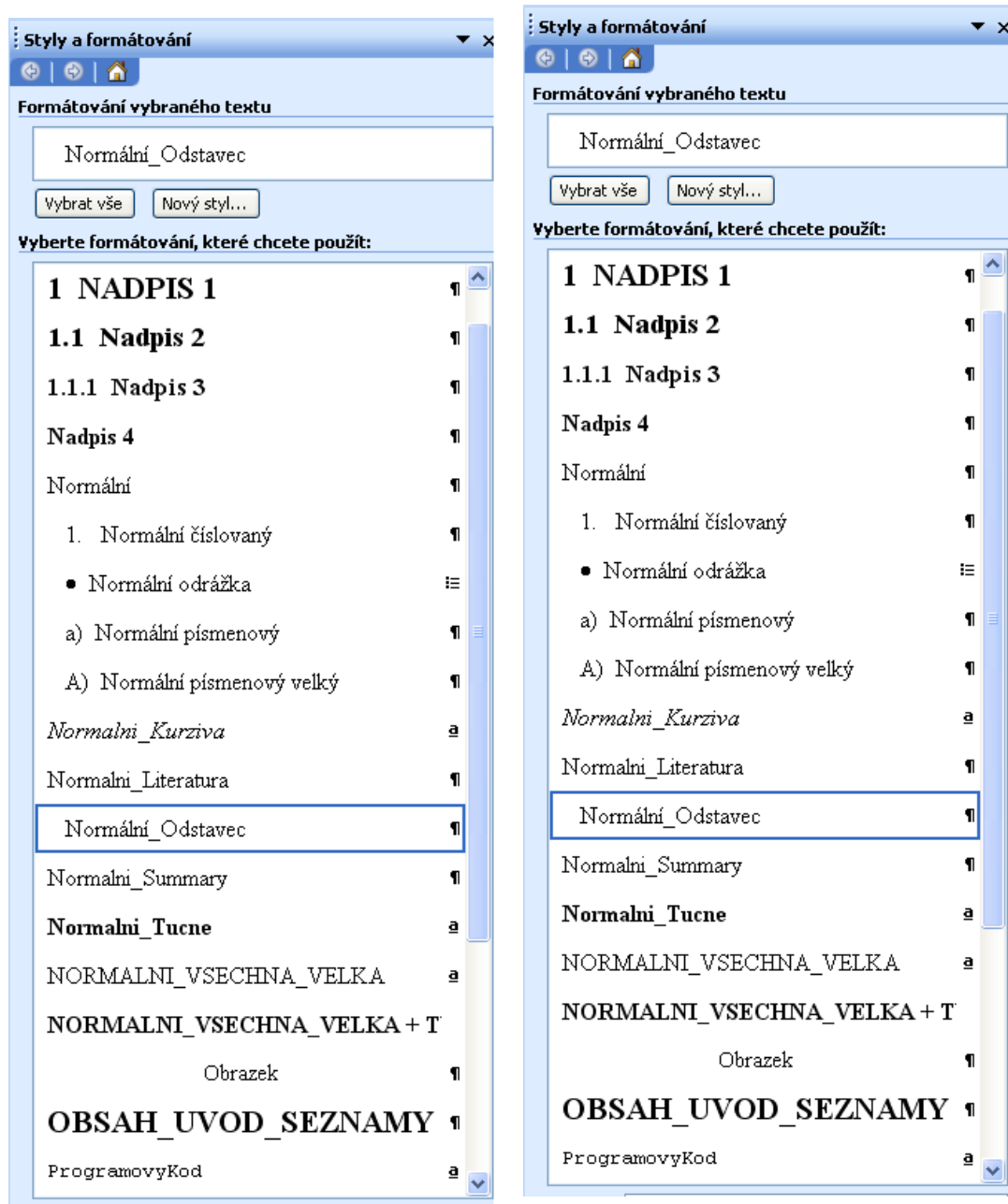
Tabulka 1: Ukázková tabulka

Team sheet		
Goalkeeper	GK	Paul Robinson
Defenders	LB	Lucus Radebe
	DC	Michael Duberry
	DC	Dominic Matteo
	RB	Didier Domi
Midfielders	MC	David Batty
	MC	Eirik Bakke
	MC	Jody Morris
Forward	FW	Jamie McMaster
Strikers	ST	Alan Smith
	ST	Mark Viduka

Odkaz na tabulku pak vytvoříme takto: (viz tab. 1).

$$c = \sqrt{a^2 + b^2} \quad (1)$$

Vzorce pak odkazujeme (viz (1)).



(a) Obsah zinku v půdě

Obrázek 1: Povrch socioekonomického (a) a fyzickogeografického (b) ukazatele

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 a komplexní analýzu dat (Momjian, 2001). Systém řízení báze dat¹ 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 i jiného formátu určeného pro ukládání dat, například tabulkového procesoru. Pro komplexní správu dat velkého projektu je však databáze více než vhodná.

Prostorová databáze, někdy také zvaná geodatabáze, není nic jiného než databáze přidaná 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 3.4.1 PostgreSQL 9.x (PostGIS) a 3.4.2 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 (Esri, 2006). Příkladem takového formátu může být vektorový Esri shapefile, Esri coverage, GML, KML, GeoJSON nebo rastrový Erdas Image a GeoTIFF. Dalším způsobem je již zmíněná databáze, do níž se vektorová data ukládají ve specifickém tvaru daném standardem OGC² 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:

- bod - POINT(0 0)
- linie - LINESTRING(0 0, 1 1, 1 2)
- polygon - POLYGON ((0 0, 4 0, 4 0, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1))
- série bodů - MULTIPOINT((0 0), (1 2))
- série linií - MULTILINESTRING((0 0, 1 1, 1 2), (2 3, 3 2, 5 4))

¹V anglickém originále Database Management System (DBMS)

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

- geometrická kolekce, která může obsahovat různé geoprvky (body, linie i polygony) - `GEOMETRYCOLLECTION(POINT(2 3),LINESTRING(2 3,3 4))`³

První slovo specifikace určuje druh prvku (point, linestring, polygon, multi-point, ...), následují v závorce vypsané souřadnice lomových bodů. Za tím ještě může následovat volitelný parametr kód souřadnicového systému.

Hodnoty lze dále vkládat přes Well-Known Binary (WKB) nebo Well-Known Text (WKT) reprezentaci. PostGIS funkce pro vkládání geometrie vypadá následovně:

- `ST_AsBinary(geometry)` pro bitový zápis WKB
- `ST_AsText(geometry)` pro WKT text

Příklad uložení linie do databáze s jedním lomovým bodem v souřadnicovém systému WGS84:

```
(LINESTRING(15.96 50.84, 17.29 49.64, 18.27 49.80), 4326) (1)
```

3.1 Vymezení pojmů

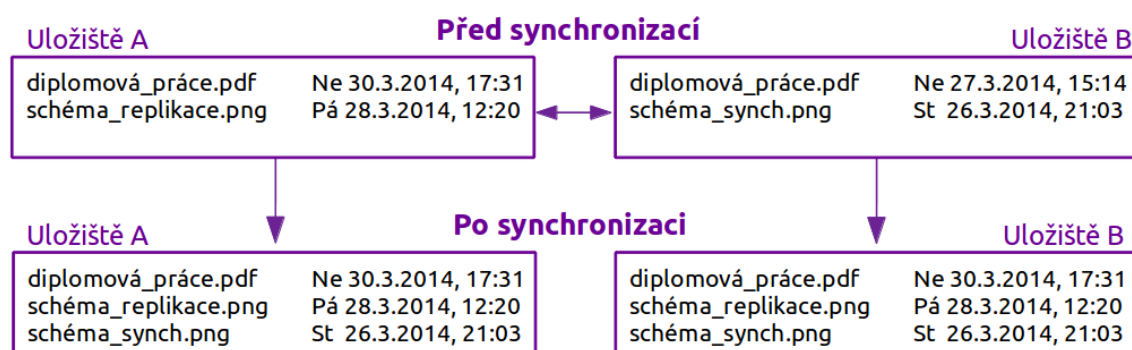
Pro lepší porozumění textu této práce je potřeba definovat pojmy replikace, synchronizace a verzování, včetně popisu toho, jak jsou dané pojmy chápány v produktech ArcGIS. Je vhodné upozornit, že výše zmíněné procesy jsou v literatuře často chápány lehce odlišně. Některé zdroje pojmy replikace a synchronizace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

Všechny dotyčné pojmy úzce souvisí se zálohováním dat, tedy kopírováním dat mezi dvěma a více uložišti. 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 konkrétním způsobem provedení zálohy, či přesným důvodem kopírování dat.

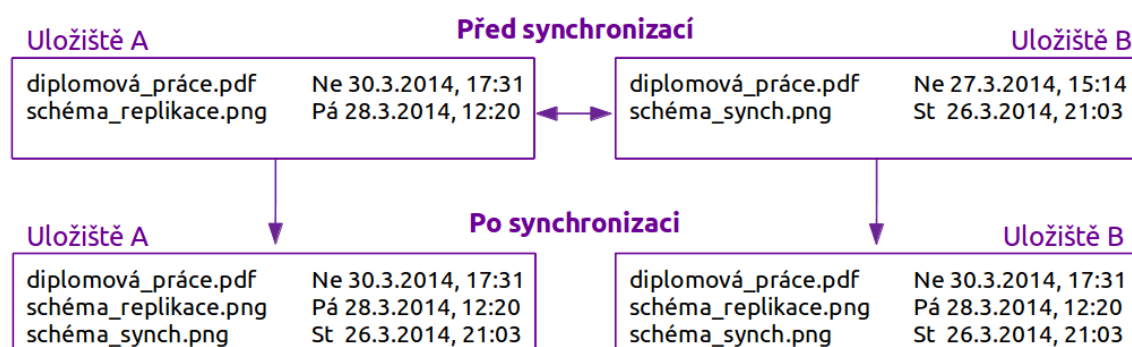
Z mého pohledu je synchronizace nadmnožinou replikace. V případě, že existují dva datové zdroje a v jeden okamžik se rozhodneme, že chceme tyto dvě složky sjednotit, poté je možno mluvit o synchronizaci souborů či datových složek. Soubor, který se podle názvu nachází ve složce A a zároveň se nenachází 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

³Zdroj http://postgis.net/docs/manual-2.1/using_postgis_dbmanagement.html#RefObject

označit v okamžiku, kdy existují nejméně dva datové zdroje a smyslem synchronizace je porovnat tato uložení a dostat je do stejného stavu. To může například přispět snazší spolupráci více uživatelů nad stejnými daty nebo uživateli, který pracuje na více počítačích.



Obrázek 2: Příklad obousměrné synchronizace dat mezi dvěma datovými uloženími



Obrázek 3: Příklad obousměrné synchronizace dat mezi dvěma datovými uloženími

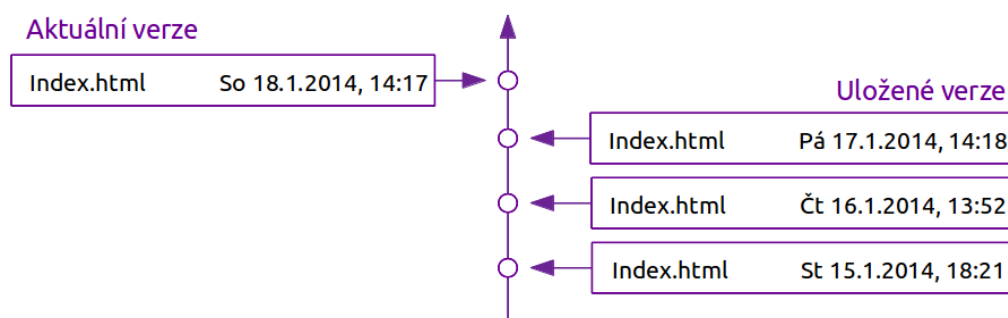
Replikace naopak, podle mého názoru, začíná s daty existujícími pouze na jednom uložení. Často je tento proces používán právě ve spojitosti s databázemi, kdy je kopie dat (také replika) tvořena z důvodu snížení zátěže serveru, či ochrany dat. V případě, že je tato kopie již vytvořena, je poté možno mluvit 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 uložení na druhé a nikoliv opačně, nebo oboustranně, kdy se datové zdroje kopírují navzájem mezi sebou.

Specifickým způsobem zálohy dat je verzování, kdy se data na záložním datovém uložení nepřepisují, ale systematicky ukládají v takzvaných verzích tak, aby se uží-

vatel 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. To, co je zde popsáno jako verzování, se v produktech ArcGIS nazývá archivování dat (Law, 2008).

Verzování může probíhat ručně, poloautomatizovaně či plně automatizovaně díky speciálním nástrojům pro správu verzí. Oblíbeným verzovací systémem programátorů je GIT⁴, 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, c2009). GIT je vhodný zejména pro textové soubory, protože dokáže analyzovat části textu, či programového kódu a zvýraznit části, které se změnily.



Obrázek 4: Příklad verzování souboru

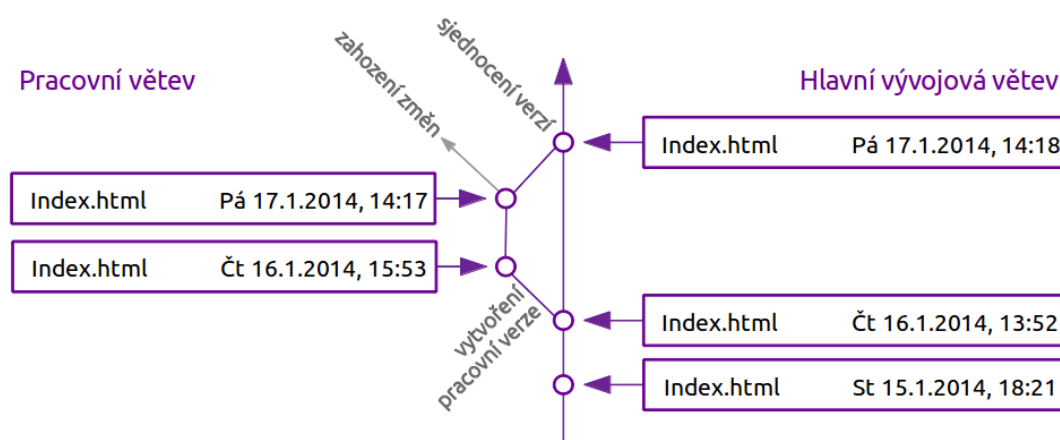
Samotná databáze verzování dat neumožňuje. Nejsnazším způsobem, jak získat verzi dat, je dump, tedy export databáze do souboru. V MS SQL Serveru je tento proces nazýván Snapshot, tedy snímek databáze nebo také snímková replikace. Takový soubor se poté může verzovat podobným způsobem jako jakýkoliv jiný 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 gitu a nese název GeoGIT. Umožňuje uživatelům zachovávat změny v souborech shapefile, SpatialLite a z databáze PostGIS (PostgreSQL). Umožňuje, tak jako git, uchovávat historii prostorových dat, či vrátit se k předchozí verzi.

Verzování může být chápáno také jako vytvoření pracovní verze. V případě, že programový kód či data jsou plno funkční či aktuální, ale je potřeba je testovat či jinak měnit, pak je vhodné vytvořit tzn. pracovní verzi, aby nedošlo k poškození té správné. Pracovní verze je kopie aktuálního stavu, na které je možno pracovat a zkoušet. V případě, že práce nedopadne podle představ, je možno změny zahodit, pokud je tomu

⁴Více na <http://git-scm.com/>

naopak, je možno pracovní verzi sjednotit s platnou verzí. Tento způsob verzování umožňuje GIT i GeoGIT a takto chápe pojem verzování i společnost Esri.



Obrázek 5: Příklad verzování souboru s použitím pracovní větve

3.2 Replikace

Replikace je proces, u kterého jsou data a databázové objekty kopírované z jednoho databázového serveru na druhý a poté synchronizovány 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).

Mnohé moderní aplikace se musí zabývat velkým počtem přístupů do databáze, což může v některých případech způsobovat problémy. Buď je server přetížen počtem připojení a data tedy přicházejí k uživateli pomalu, nebo dokonce úplně vypadne.

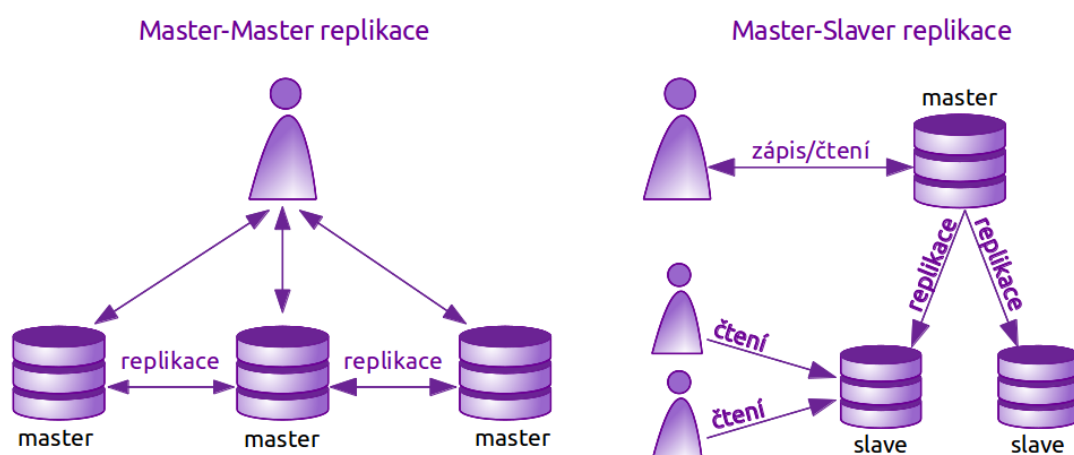
Mezi časté důvody použití databázové replikace tedy patří zajištění dostupnosti dat⁵, resp. snížení pravděpodobnosti, že data nebudou dostupná, což může být způsobeno již zmíněným výpadkem serveru nebo například fyzickou ztrátou dat (Obe a Hsu, 2012). Další důvodem je rozložení zátěže přístupů do databáze mezi více serverů, takže nebude docházet ke zpomalení výkonu hlavního serveru ani k situaci, že data nebudou dostupná kvůli jeho výpadku (Bell et al., 2010). Databáze je často zálohovaná, například skriptem dump a i to může server zpomalit. Vhodným řešením je tedy nejdříve vytvořit kopii dat na jiný datový server a až poté proces zálohování spustit.

Všechny databáze zapojené do procesu replikace jsou v odborné literatuře nazý-

⁵Anglicky High Availability

vané uzly, v angličtině node. Tyto uzly dohromady tvoří replikační cluster⁶. Při správně nastavené replikaci, by v clusteru nikdy neměly být méně než 3 uzly. Může se totiž stát, že vypadne jeden ze dvou uzlů, čímž dojde, ikdyž jen na krátkou chvíli, k situaci, že data nebudou v daný okamžik zálohovaná.

Uzly v replikačním clusteru mohou mít jednu ze dvou základních rolí, nejčastěji nazývaných Master a Slave. Master server nebo pouze Master je server, který poskytuje data k replikaci, má práva na čtení i zápis a probíhají tedy na něm veškeré aktualizace. Je možno se setkat také s pojmenováním 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, který je dostupný vždy jen pro čtení dat, se data a aktualizace kopírují, není však možné na něj změny zapisovat (Riggs a Krosing, 2010).



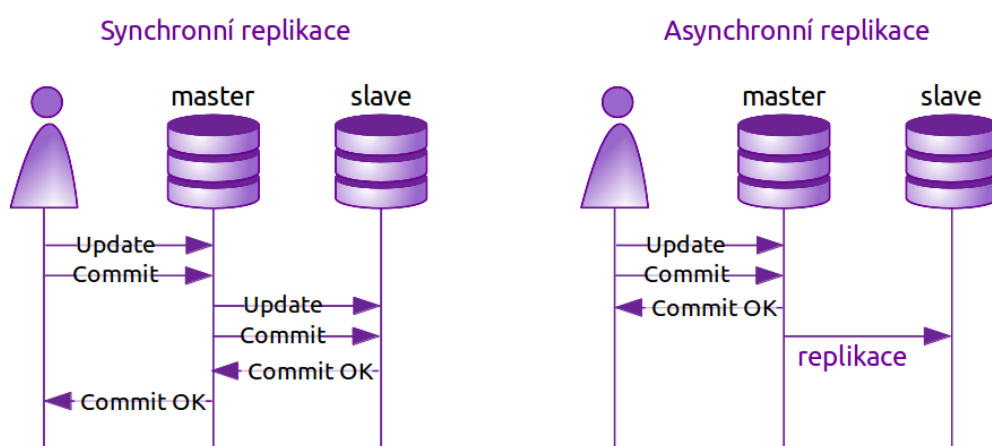
Obrázek 6: Srovnání Master-Master a Master-Slave replikace

Podle počtu Master a Slave serverů v replikačním clusteru, se rozlišuje zda se jedná o jednosměrnou nebo obousměrnou replikaci. Tzv. Master-Master replikace umožňuje zapisovat do všech uzlů v replikačním clusteru, což může být praktické například při použití databáze offline (viz obr. 8). Změny se tedy synchronizují mezi všemi databázovými uzly. Tento způsob však nese značné komplikace, je potřeba řešit konflikty změn ve stejných datech a je relativně náročný na údržbu. Tato práce se zabývá použitím druhé způsobu, tzv Master-Slave replikace. Tato replikace používá vždy jen jeden Master server v clusteru a dva a více Slave servery. Kopie dat tedy probíhá jednosměrně, vždy z Master na Slave servery. Podle Bella (2010) mají moderní aplikace často více čtenářů než zapisovatelů, proto je zbytečné, aby se všichni čtenáři

⁶Volně přeloženo skupina serveru zapojených do replikace

připojovali na stejnou databázi jako zapisovatelé a zpomalovali tím jejich práci (Bell et al., 2010). Z toho důvodu je tedy použití Master-Slave replikace více než vhodné.

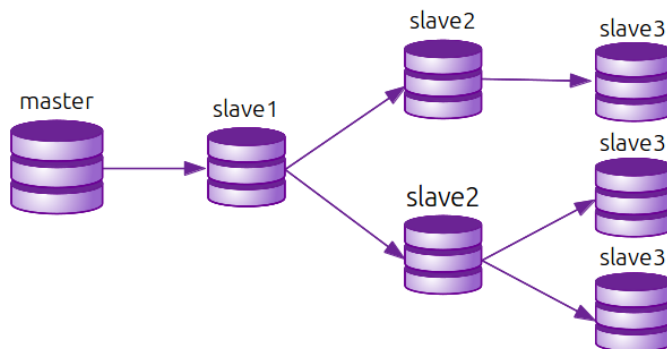
Při návrhu je potřeba se zamyslet také nad způsobem replikace, zda bude synchronní či asynchronní. Synchronní replikace zajišťuje, že na Master serveru nikdy neproběhne nová transakce, dokud se poslední transakce neprovede na Slave serveru (Böszörmenyi a Schönig, c2013). Tento přístup zajistí, že žádná data nebudou v průběhu transakce ztracena. V některých případech tento způsob může zbytečně zpomalit rychlost přístupu do databáze, protože je nutno čekat na každou nedokončenou transakci. Naopak při bankovních transakcích, kde je potřeba, aby všechny operace proběhly na všech stranách, je tento způsob nezbytný. Druhým způsobem je asynchronní replikace, při které se nová data mohou zapisovat na Master server, přestože ještě nedošlo k replikaci stávajících dat na Slave server (Obe a Hsu, 2012).



Obrázek 7: Rozdíl mezi synchronní a asynchronní replikací

Replikace v PostgreSQL umožňuje plnou kopii dat z databáze i pouze výběr některých tabulek. Více o možnostech a způsobech nastavení replikace v kapitole PRAKTICKÁ ČÁST :)

Každý databázový server (myšleno SŘDB) si volí terminologii a konkrétní nastavení mírně odlišně. Tato kapitola se snaží popsat chápání replikace co v největší míře obecně s ohledem na použití tohoto pojmu v PostgreSQL. Zcela jinou terminologii, ikdyž založenou na stejných principech, zavádí MS SQL Server, který používá pojmy transakční replikace pro Master-Slave replikace a slučovací replikaci pro Master-Master replikaci.



Obrázek 8: Ukázka kaskádové replikace

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 1969⁷.

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 z katalogu (kolekce geografických datových sad), map a sad nástrojů pro geografické analýzy.

Esri vytváří integrovanou sadu softwarových produktů ArcGIS, které poskytují nástroje na kompletní správu GIS a přizpůsobují 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).

Základními produkty⁸ jsou stolní systémy ArcGIS for Desktop ve verzích Basic, Standard, Advanced⁹, 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 ArcPad, ArcGIS for Windows Mobile a ArcGIS for Smartphone and Tablet, je určený

⁷Více info <http://www.esri.com/about-esri/history>

⁸Názvy 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/>.

⁹Zdroj <http://www.esri.com/software/arcgis/about/gis-for-me>

především pro práci v terénu. A v neposlední řadě verze dostupná skrze internet ArcGIS Online. K tomu všemu Esri přidává velké množství extenzí a další verzí¹⁰.

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			

Obrázek 9: Verze programu ArcGIS platné od verze 10.1.

Dle Law (2008) je nativním formátem produktů ArcGIS geodatabáze a jsou rozlišovány tři druhy geodatabáze. Ani v jednom případě se však nejedná o databázi v pravém slova smyslu, tak jako ji chápáme v kapitole 3.4.1 a 3.4.2. V každém případě však tyto způsoby 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.

¹⁰ Kompletní seznam na oficiálních webových stránkách Esri <http://www.esri.com/products> nebo <http://www.arcdata.cz/produkty-a-sluzby/software/arcgis/>

databáze		souborová .gdb ¹	personální .mdb ¹
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 ²	1 ²
počet čtenářů		více než 1 ²	více než 1 ²
replikace	typ replikace	souborová replikace	souborová replikace
	master server	ne ¹	ne ¹
	slave server	ano	ano
verzování		ne	ne
velikostní limity		1TB pro každý dataset ²	2GB ²

Obrázek 10: Přehled rozdílů personální a souborové geodatabáze používané programem ArcGIS

¹<http://www.esri.com/software/arcgis/geodatabase/single-user-geodatabase>

²<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/003n00000007000000>

Tato práce se více zaměřuje na třetí typ, technologii ArcSDE, kterou v některých materiálech nazývají “geodatabáze ArcSDE”. Nejedná se o geodatabázi, ale spíše o zprostředkovatele komunikace mezi programem ArcGIS a databázovým serverem. 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. Touto technologií se více bude zabývat kapitola 3.4.3 ArcSDE geodatabase.

3.4 Použité programové prostředky

3.4.1 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 před 20 lety. 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 a 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. Je zejména vhodný pro malá území. 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 (OpenGeo, 2012b).

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 a Hsu, 2011). PostGIS 1.5. obsahovala přes 800 funkcí, typů a

prostorových indexů (Obe a Hsu, 2012). Aktuální verze PostGIS¹¹ je 2.1.

PostgreSQL podporuje replikaci i synchronizaci bez nutnosti další instalace.

Od verze ArcGIS 9.3. je PostgreSQL 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 SP1¹² je to PostgreSQL 9.1.3 a PostGIS 2.0 (OSGeo, 2013)¹³. Databáze PostgreSQL se dá v ArcGIS produktech použít dvojím způsobem. Buď jen jako uložisko 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 PostgreSQL¹⁴ 9.1.4 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 (nebere 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

¹¹Aktuálně na <http://postgis.refractory.net/>

¹²Service Pack 1

¹³Zdroj 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>

¹⁴Více na <http://www.postgresql.org/>

zakoupenou licenci 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 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). Stejně jako u PostgreSQL platí, že si uživatel může zvolit, zda použije datový typ, který je součástí ArcSDE, nebo ten, který je implementován do SQL Serveru.

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. Využívá jako své nativní datové struktury standard konsorcia OGC 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). Každá verze má jiné parametry a umožňuje různou úroveň editace (viz obr. 11).

databáze		ArcSDE		
		Desktop ¹	Workgroup ¹	Enterprise ¹
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

Obrázek 11: Přehled verzí ArcSDE, jejich parametrů a možností

¹ zdroj <http://www.esri.com/software/arcgis/geodatabase/multi-user-geodatabase>

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 3.4.2 Microsoft SQL Server Express 2008, SQL Server Express je možný použít v replikačním clusteru pouze jako slave server. Vzhledem k tomu, že proces replikace je implementován při do ArcObjects a ArcSDE, nezáleží na konkrétním databázovém systému (Law, 2008).

4 VÝSLEDKY

5 DISKUZE

6 ZÁVĚR

LITERATURA

- BELL, C., KINDAHL, M., THALMANN, L. *MySQL high availability*. Vyd. 1. Sebastopol, CA: O'Reilly Media, Inc, 2010. ISBN 978-059-6807-306.
- BöSZÖRMENYI, Z., SCHÖNIG, H.-J. *PostgreSQL replication: understand basic replication concepts and efficiently replicate PostgreSQL using high-end techniques to protect your data and run your server without interruptions*. Birmingham: Packt Publishing, c2013. ISBN 978-1-84951-672-3.
- CHACON, S. *Pro Git*. Edice CZ.NIC. Praha: CZ.NIC, c2009. ISBN 978-80-904248-1-4.
- CONNOLLY, T. *Database Systems: A Practical Approach to Design, Implementation, and Management*. Vyd. 4. Harlow: Addison-Wesley, 2005. ISBN 03-212-1025-5.
- ESRI. *ArcGIS 9: Co je ArcGIS 9.2?* United States: ESRI Press, US, 2006. ISBN 15-894-8166-6.
- ESRI. A quick tour of working with databases in arcgis. *ArcGIS Help 10.1 [online]*, 2013a. Dostupné z: http://resources.arcgis.com/en/help/main/10.1/index.html#/A_quick_tour_of_working_with_databases_in_ArcGIS/019v00000008000000/.
- ESRI. Preparing data for replication. *ArcGIS Help 10.1 [online]*, 2013b. Dostupné z: http://resources.arcgis.com/en/help/main/10.1/index.html#/Preparing_data_for_replication/003n000000z5000000/.
- LAW, D. Enterprise geodatabase 101: A review of design and key features for gis managers and database administrators. *Esri: Understanding our world. [online]*, 2008. Dostupné z: http://www.esri.com/news/arcuser/0408/entergdb_101.html.
- LEITER, C. *Beginning Microsoft SQL server 2008 administration*. Indianapolis, IN: Wiley Pub., 2009. ISBN 978-047-0440-919.
- MICROSOFT. SQL server - replication. *Microsoft [online]*, 2013. 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).
- MOMJIAN, B. *PostgreSQL: introduction and concepts*. Boston, MA: Addison-Wesley, 2001. ISBN 02-017-0331-9.
- OBE, R., HSU, L. *PostGIS in action*. London: Pearson Education [distributor], 2011. ISBN 19-351-8226-9.

- OBE, R., HSU, L. *Postgresql: Up and Running*. Sebastopol, CA: O'Reilly, 2012. ISBN 978-144-9326-333.
- OPENGEO. Introduction to postgis [online]. *Section 17: Geography*, 2012b. Dostupné z: <http://workshops.opengeo.org/stack-intro/openlayers.html>.
- OPPEL, A. J. *Databases: a beginner's guide*. New York: McGraw-Hill, 2009. ISBN 00-716-0846-X.
- OSGEO. Postgis and arcsde/arcgis articles. *PostGIS Tracker and Wiki [online]*, 2013. Dostupné z: <http://trac.osgeo.org/postgis/wiki/UsersWikiPostgisarcgis>.
- POSTGRESQL. Faq. *PostgreSQL wiki [online]*, 2012. Dostupné z: <http://wiki.postgresql.org/wiki/FAQ>.
- RIGGS, S., KROSING, H. *PostgreSQL 9 administration cookbook: solve real-world PostgreSQL problems with over 100 simple, yet incredibly effective recipes*. Birmingham: Packt Publishing, 2010. ISBN 978-1-849510-28-8.
- WHALEN, E. a. k. *Microsoft SQL Server 2005: velký průvodce administrátora*. Vyd. 1. / *Edice Administrace (Computer Press)*. Brno: Computer Press, 2008. ISBN 978-80-251-1949-5.
- ČINČURA, J. MS SQL 2008 – prostorová data poprvé. *Databázový svět [online]*, 2009. Dostupné z: <http://www.dbsvet.cz/view.php?cisloclanku=2009101201>.

SUMMARY

There is summary of all aims, methods and results in this chapter. Summary is not only translation of chapter Závěr. There is more information from chapters Cíle, Výsledky and Diskuze. Number of pages of Summary chapter is two at least. The style is Normalni Summary. Language is set to Angličina(Velká Británie) for automatic spell check. Do not use language Angličtina(USA).

PŘÍLOHY

SEZNAM PŘÍLOH

Volné přílohy

Příloha 1 CD

Popis sktruktury CD

Adresáře a soubory:

- složka se skripty
- web - webová stránky jako doplněk k diplomové práci
- Solanska_dp.pdf - text diplomové práce