

— Posudek vedoucího diplomové práce —

Název: Processing of vector data using distributed database systems in GIS
Student: Bc. Matěj Krejčí
Vedoucí: Ing. Martin Landa, Ph.D.
Fakulta: Fakulta stavební ČVUT v Praze
Katedra: Katedra geomatiky
Oponent: Ing. Jan Pytel, Ph.D.
Pracoviště opONENTA: Katedra geomatiky, FSv ČVUT v Praze

Zadání původně vzešlo z bakalářské práce, kterou student zpracoval v roce 2014 na téma „Analýza a vizualizace srážkových dat z mikrovlnných telekomunikačních spojů pomocí GIS“. Jejím praktickým výstupem byla sada nástrojů pro systém GRASS a databáze mikrovlnných spojů provozovaná na školním serveru geo102. Oba výstupy byly vytvořeny „na objednávku“ pro kolegy z katedry hydrauliky a hydrologie v rámci řešeného grantu GAČR. Jako databázové uložení byl zvolen objektově-relační databázový systém PostgreSQL. Toto řešení se postupem času s narůstajícím počtem záznamů (aktuálně přes 3 miliardy záznamů) ukázalo jako nedostačující a nevhodné. V této době jsme společně s kolegy uvažovali jak s těmito daty naložit optimálnějším způsobem, což nás vedlo přirozeně k problematice tzv. velkých dat („big data“). Na základě této úvahy vznikla předkládaná diplomová práce.

Text práce je napsán v anglickém jazyce. Důraz je kladen především na problematiku zpracování geografických dat, resp. dat s prostorovou složkou. Přestože je téma vysoce aktuální a přínosné, tak nám není známo, že by se mu věnovaly další práce na našem oboru. Autor popisuje problematiku velkých dat nejprve obecně. V další části se věnuje specifikům práce s geografickými daty, dostupných knihoven a technologii jako celku. Na začátku práce jsme uvažovali o technologiích, které budou použity pro její praktickou část. V prvotní fázi student experimentoval se sítí počítačů virtualizovaných jako kontejnery Docker. Tento postup se jevil jako příliš komplikovaný, proto byla po konzultaci s kolegou Ing. Janem Pytlem, Ph.D. zvolena technologie Google Cloud. Google navíc nabízí bezplatný program pro studenty, který umožňuje jejich technologie a servery využívat bezplatně po dobu tří měsíců. Této možnosti bylo pro zpracování práce využito.

Praktická část práce je postavena na dvou datových sadách. První navazuje na výše zmiňovanou bakalářskou práci. Jde o export z databáze mikrovlnných spojů. Zde si student vyzkoušel praktickou práci s platformou Google Cloud a importem velkých dat, která vyexportoval z lokální databáze PostgreSQL ve formátu CSV (viz příloha A). Tato datová sada má fixní prostorovou složku (umístění vysílačů se v čase nemění), dynamické jsou v čase přirozeně naměřené hodnoty. Vzhledem k tomu není tato sada vhodná pro prostorové analýzy, typicky typu agregace či zonální statistiky. Bylo tedy třeba najít takovou datovou sadu, která splňuje co do počtu prvků podmínku velkých dat a zároveň je její dynamika v čase svázána právě s prostorovou složkou popisu geoprvků. Z volně dostupných zdrojů dat padla volba na databázi OpenStreetMap. Zpracována byla datová sada pro Evropu, která má včetně historie od jejího vzniku přes 1.3 bilionů prvků. Výsledek tohoto snažení je popsán v kapitole 3.2. Zde je také demonstrováno propojení se systémem GRASS GIS. S tím souvisí zásadní praktický výstup práce, a to sada nástrojů umožňující komunikaci mezi systémy GRASS a technolo-

giemi Hadoop a Hive. Vzniklé nástroje budou dostupné pro ostatní uživatele systému GRASS v rámci tzv. Addons.

Zadání práce vzniklo ze studentovi iniciativy, byl schopen se v relativně krátkém čase zorientovat pro něj v naprosto nové problematice a rozumně ji uchopit. Na základě toho vznikl praktický výstup práce, a to nástroje pro propojení systému GRASS a technologií Hadoop, Hive. Z těchto nástrojů mohou těžit a dále je rozvíjet i další uživatelé systému GRASS. Zpracování práce považuji za naprosto odpovídající vstupním podmínkám a jako vedoucí práce jsem s jejím výsledkem spokojen.

Na základě výše uvedeného hodnotím předloženou diplomovou práci klasifikačním stupněm

— **A (výborně)** —

V Solanech dne 17. června 2016

.....
Ing. Martin Landa, Ph.D.
Fakulta stavební, ČVUT v Praze