

Posudek oponenta

Diplomová práce: Využití distribuovaných databázových systémů pro správu vektorových dat v GIS
Diplomant: Bc. Matěj Krejčí
Vedoucí: Ing. Martin Landa, Ph.D.
Oponent: Ing. Jan Pytel, Ph.D.

Žijeme v době, kdy problematika *big data* nabývá na významu téměř ve všech odvětvích, každý den jsou celosvětově vygenerována nová data o celkové velikosti 2.5 exabytů. Například jeden let letadla Boeing 787 vygeneruje 0.5 TB dat. Diplomová práce pana Matěje Krejčího tuto skutečnost reflektuje a zabývá se problematikou a zpracováním *big data* s využitím frameworku Hadoop.

Diplomová práce je rozdělena na dvě části:

teoretická část — obsahuje popis použitých frameworků a postupů

praktická část — obsahuje praktickou implementaci pojmenovanou GRASS Hadoop Framework.

Úvodní kapitola teoretické části obsahuje popis frameworku Hadoop, diplomant zde popisuje všechny obsažené moduly. Z textu je patrná architektura ukládání/získávání dat i následná práce s daty s využitím paradigmatu *MapReduce*.

Druhá kapitola obsahuje popis paralelního zpracování prostorových dat. Nejprve jsou popsány způsoby možného rozdělení prostorových dat (spatial data skew, boundary object) a indexace. Poté přichází popis a možnosti knihoven vhodných pro použití paradigmatu *MapReduce* v prostředí frameworku Hadoop.

Třetí kapitola se zabývá popisem *Google Cloud Platform*, tedy platformou firmy Google, kde je možno hostovat aplikace s možností využití řady frameworků včetně frameworku Hadoop. V kapitole je stručný návod, jak vytvořit v prostředí *Google Cloud Platform* podvozek pro budoucí projekt.

Závěrečná kapitola teoretické části krátce popisuje projekt GRASS.

Praktická část je rozdělena na čtyři kapitoly. V první kapitole je popsáno nastavení síťové infrastruktury a frameworku Hadoop, využitých v následujících kapitolách.

Druhá kapitola popisuje diplomantův framework *GRASS Hadoop framework*. Framework umožňující interakci mezi daty uloženými v prostředí frameworku Hadoop a GRASS GIS. Kapitola obsahuje funkcionalitu frameworku a způsob implementace. Právě v této kapitole diplomant zužitkoval znalosti a zkušenosti popsané v předchozích kapitolách.

Třetí kapitola ukazuje framework *GRASS Hadoop framework* během praktického využití.

Diplomová práce je napsána čtivou formou a obsahově odpovídá zkušenostem a časovým možnostem studenta magisterského studia.

Dotazy k diplomové práci

- Jaké další frameworky typu Hadoop existují?
- Pokud je možné Hadoop provozovat na Google cloudu, jaké jsou další možnosti, kde lze řešení na Hadoopu v cloudu používat? Jaké další Cloud computing services diplomant zná?
- Proč byl zvolen pro uložení *big data* právě framework Hadoop?

- Jaké jsou výhody a nevýhody NoSQL databází vs “typické” relační SQL databáze?

Předloženou diplomovou práci považuji za práci, jenž odpovídá zkušenostem a možnostem studenta magisterského studia a hodnotím ji proto známkou

“A” výborně

a domnívám se, že předložená práce si zaslouží ocenit pochvalou za vynikající zpracování.

V Praze 21. 6. 2016

Ing. Jan Pytel, Ph.D.