

Posudek oponenta diplomové práce

Student: Krobot Pavel, Ing.
Téma: Distribuované zpracování dat o IP tocích (id 17592)
Oponent: Kořenek Jan, Ing., Ph.D., UPSY FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Cílem práce bylo navrhnout vhodný způsob distribuovaného zpracování dat o IP tocích. Student analyzoval dostupné prostředí pro zpracování velkých dat a na základě výsledků navrhl a implementoval vlastní řešení. Rozsahem provedených experimentů, náročností řešené problematiky i složitostí implementace považuji práci za velmi náročnou.
- 2. Splnění požadavků zadání** **zadání splněno**
Všechny body zadání byly splněny.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Rozsah technické zprávy odpovídá požadavkům kladeným na diplomovou práci.
- 4. Prezentací úroveň předložené práce** **80 b. (B)**
Technická zpráva je napsána přehledně a kapitoly jsou uspořádány v logickém sledu. Vytkl bych pouze, že text je místy více strukturován do odrážek, tabulek a grafů. Zejména v části, kde jsou definovány požadavky na distribuovaný systém a popis implementace.
- 5. Formální úprava technické zprávy** **70 b. (C)**
Po jazykové stránce je práce velmi dobře zpracována. Vytvořený text má charakter odborného textu, je dobře čitelný a snadno pochopitelný. Po typografické stránce je práce velmi dobrá. Vytkl bych pouze několik překlepů, které mohly být odstraněny kontrolou pravopisu. Obrázek architektury v kapitole 4 by měl být ve vyšším rozlišení, tabulka 4.5 by měla být vložena jako tabulka. I přes uvedené nedostatky je formální úprava technické zprávy dobrá.
- 6. Práce s literaturou** **85 b. (B)**
Student se musel seznámit se systémy pro zpracování velkých dat. Studijní prameny jsou voleny vhodným způsobem a jsou v textu správným způsobem citovány.
- 7. Realizační výstup** **90 b. (A)**
Realizačním výstupem je funkční program DistDump, který umožňuje zpracování velkých dat o IP tocích v distribuovaném prostředí.
- 8. Využitelnost výsledků**
Práce přináší nový způsob distribuovaného zpracování IP toků. Byla vytvořena funkční implementace distribuovaného kolektoru, která ukazuje možnosti distribuce zátěže na více počítačů. Vytvořená implementace má potenciál stát se základem systému pro zpracování velkých dat o IP tocích. Přínosem jsou také výsledky experimentů, které ukazují využití současných nástrojů pro zpracování velkých dat v řešené problematice.
- 9. Otázky k obhajobě**
 - V práci je uvedeno srovnání navrženého řešení s programem NfDump a ukázána škálovatelnost na čtyřech počítačích. Jaké vidíte u navrženého řešení limity ve škálovatelnosti? Kde by vzniklo úzké místo navrženého systému?
 - Plánuje se nasazení vytvořené implementace DistDump v produkční síti?
- 10. Souhrnné hodnocení** **85 b. velmi dobře (B)**
Celkově hodnotím diplomovou práci jako velice zdařilou. Student musel nastudovat oblast zpracování velkých dat pomocí Map/Reduce přístupů, seznámit se s dostupnými nástroji a provést řadu experimentů, které ukazují na nízkou využitelnost existujících řešení. Proto student navrhl vlastní řešení, které má pro zpracování dat o IP tocích výrazně lepší parametry, než dostupné nástroje. Práce přináší nové poznatky i novou implementaci, která se může stát základem distribuovaného kolektoru síťových toků. Proto navrhuji hodnocením stupněm A (výborně).

V Brně dne: 12. června 2015

.....
podpis

