

Posudek na diplomovou práci
Improving Efficiency of HEP Applications

Jiří Horký

Práce se zabývá analýzou přístupů k datům během vyhodnocování experimentů provedených na Velkém hadronovém urychlovači a porovnáním efektivity vybraných distribuovaných souborových systémů.

V první kapitole autor popisuje jakým způsobem probíhá distribuce dat mezi jednotlivými výpočetními centry a spolu s druhou kapitolou, ve které popisuje strukturu jednotlivých úloh, tvoří odborný základ problematiky, kterou se zabývá zbytek práce. Text je i přes množství informací srozumitelný.

Třetí kapitola popisuje možné implementace profilování aplikací. Autor na základě analýzy výhod a nevýhod jednotlivých způsobů vybírá nástroj strace a ve čtvrté kapitole popisuje implementaci vlastního programu IOprofiler, který výstup programu strace zpracovává a vizualizuje. Pomocí IOprofileru pak v páté kapitole analyzuje přístupy jednotlivých úloh k souborům z hlediska požadavků na vyhledávání na discích.

V šesté kapitole autor popisuje druhý program, IOreplay, který na základě výstupů z programu strace dovede simulovat chování úlohy z hlediska vstupů/výstupů. Autor rovněž zdůvodňuje, v čem spočívají výhody použití programu IOreplay pro analýzu efektivity souborových systémů před testováním na reálně spuštěných úlohách. Nakonec ukazuje, že tento přístup poskytuje použitelné výsledky pro účely testování.

Sedmá kapitola obsahuje popis metodiky měření experimentů a samotná měření pro souborové systémy Lustre, GPFS a HDFS. Rovněž popisuje důležité vlastnosti těchto systémů. Měření NFS4.1 chybí, což autor zdůvodňuje tím, že systém je ve vývoji a nepodařilo se mu systém pro testování zprovoznit. Nechybí ani vzájemné porovnání jednotlivých systémů z hlediska výkonu a zátěže sítě.

Kapitola osmá seznamuje čtenáře s pracemi, které se zabývají podobnou tematikou, a poslední kapitola celou práci shrnuje uvedením dosažených výsledků a plánů do budoucna.

Pozitivně hodnotím, že součástí je dvojice programů IOprofiler a IOreplay, které jsou volně dostupné a které jsou použitelné pro analýzu jiných libovolných aplikací, než těch, které jsou zmíněné v práci.

Za největší výtku považuji absenci testování systému NFS4.1, který je zmíněn v abstraktu. Na druhou stranu se autor snažil (bohužel neúspěšně) tento nedostatek vyřešit přímo s autory systému. Dále mám drobné výhrady ke grafickému provedení obrázků 2.2 a 2.3, kde černá barva na pozadí je velmi netradiční. Navíc osa x má u obou grafů trochu matoucí popisek. Grafy 7.7 a 7.8 mi přijdou mírně nekonzistentní, ačkoliv ukazují prakticky totéž (jednou porovnání výkonů, podruhé síťového zatížení), je zvolena pokaždé jiná reprezentace.

Celá práce je zpracována kvalitně, je napsána srozumitelně, bez větších stylistických, gramatických či typografických chyb. Strukturální členění je rovněž v pořádku.

Doporučuji tuto diplomovou práci k obhajobě.

14.1.2011

Mgr. Zbyněk Falt

KSI, MFF UK