

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Pavel Marek

Název práce ALTREP Data Representation in FastR

Rok odevzdání 2020

Studijní program Informatika **Studijní obor** Softwarové systémy

Autor posudku Vojtěch Horký **Role** oponent

Pracoviště Katedra distribuovaných a spolehlivých systémů

Text posudku:

Cílem práce je implementace ALTREP API programovacího jazyka R do FastR (implementace R nad GraalVM). ALTREP API by mělo umožnit zrychlit rozšiřující balíčky pro R.

Práce popisuje ALTREP rozhraní a současný stav implementace a podpory jak v GNU R (jakožto standardní implementaci) tak ve FastR. Následuje popis implementačních rozhodnutí, vlastní testy a stručné zhodnocení výkonu dodané implementace. Práce uzavírá výhledem na možná rozšíření implementace. Vlastní implementaci je v době psaní posudku připravena jako *Pull-request* do hlavní větve FastR.

Práce je poměrně přehledně strukturována, vlastní text je ale poměrně náročný kvůli velké hustotě informací. I tak je ale vcelku dobře čitelný.

Na implementaci je nutné zvláště ocenit široký záběr, kdy autor musel zvládnout nejen výše zmíněné technologie, ale navíc i implementaci GNU R a také nativní rozhraní mezi jazyky R a C a též mezi Javou a C (včetně Sulongu). Stejně tak je potřeba zmínit přípravu poměrně rozsáhlého testovacího balíčku, který funguje de facto jako test kompatibility pro ALTREP.

Text práce vyžaduje poměrně hluboké znalosti všech použitých technologií (R, FastR, GraalVM a Truffle) a v některých pasážích nahrazuje i (neexistující) dokumentaci existujícího kódu.

Z hlediska organizace textu do kapitol by některé části kapitoly třetí (*Design and implementation*) mohly být přesunuty do předchozí kapitoly a tím ponechat tuto kapitolu více zaměřenou na vlastní návrh. V současném uspořádání jsou některá rozhodnutí „utopená“ v technickém popisu stávajícího stavu.

Vzhledem k tomu, že „projekt ALTREP“ byl poprvé představen teprve před třemi roky (a to nikoliv ve smyslu hotové implementace), není příliš s čím srovnávat. Překvapivé je ale kompletní vynechání větve Microsoft R Open v celém textu (ačkoliv není důvod předpokládat rozdílný výkon v souvislosti s ALTREPem).

Vlastní implementace je poměrně rozsáhlá, *patch* vůči hlavní větvi má cca 10 000 přidanych řádků. Je nutné zmínit, že autorovi se podařilo udržet svojí větev aktualizovanou, takže otevřený *Pull-request* je bez konfliktů. Kvalita zdrojového je rozumná, k dispozici je i poměrně velké množství testů.

Pozitivně je též nutno hodnotit existenci Docker image, pomocí kterého bylo možné práci prakticky vyzkoušet bez nutnosti sestavovat všechny podpůrné nástroje.

Drobnou slabinou práce je změření výkonu dodané implementace. Implementace není bohužel ve fázi, kdy by bylo možné změřit výkon na větším balíčku (např. `vroom`) a tudíž je výkon měřen na microbenchmarkách (což ale samo o sobě nevádí). Dodaná data ale obsahují výsledky z jediného běhu, což neposkytne informace o variabilitě dat (např. kvůli JITu nebo časování GC); chybí i informace o nastavení GC (byť by mělo mít minimální vliv) a použitém hardwaru a softwaru (kvůli reprodukovatelnosti).

Vlastní benchmarky pak jsou kalibrovány směrem k podobnému skóre kvůli lepší porovnatelnosti. Otázka je, nakolik je souhrnné porovnání vůbec vypovídající a jestli dává smysl. Stejně tak volba aritmetického průměru je trochu sporná, protože je silně zatížená jediným *outlierm* a harmonický průměr by zde dával větší smysl (a kvalitu práce by určitě neovlivnilo to, jestli je souhrnné skóre lepší nebo horší vůči GNU R).

Na druhou stranu je nutné ocenit rozsáhlé okomentování nejviditelnějších výsledků.

Celkově je práce na velmi dobré úrovni. Na první pohled se cíl práce jeví poměrně jednoduchý: doplnit implementaci „jen“ cca 10 funkcí. Jeho splnění ale vyžaduje hluboké technické znalosti několika programovacích jazyků (Java, C/C++, R) i netriviálních technologií (Truffle, GraalVM či Sulong). Autor tyto znalosti dokazuje jak v textu práce tak i dodaným zdrojovým kódem hotové implementace.

Práci proto doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 28. srpna 2020

Podpis: