

Izzivi znanstvene programske opreme v superračunalništvu

dr. Alja Prah, Institut »Jožef Stefan«

Med 10. in 13. junijem 2024 je v avstrijskem Grundlsee potekalo avstrijsko-slovensko srečanje HPC. Sodelavci »Instituta Jožef Stefan«, kot tudi ostali člani konzorcija SLING, smo se srečanja v velikem številu aktivno udeležili, med drugim tudi s prispevkom *Overcoming challenges in building scientific software for HPC* oz. Premagovanje ovir pri prevajanju znanstvenih programov na superračunalnikih.

Znanstvena programska oprema je za superračunalnike seveda nepogrešljiva, saj znanstvenikom omogoča raziskovanje kompleksnih raziskovalnih problemov z uporabo ogromnih količin naprednih računskih virov. Med takšno programsko opremo najdemo širok nabor programov – npr. za poganjanje simulacij molekulske dinamike, analizo podatkov, izgradnjo jezikovnih modelov, strojno učenje –, ki zahtevajo nezanemarljivo računsko moč in uporabljajo napredne algoritme. Zaradi tega so velikega pomena pri izvajanju programske opreme na superračunalnikih tudi učinkovita paralelizacija oz. vzporedno izvajanje nalog, optimizacija za različne arhitekture in sposobnost upravljanja z veliko količino podatkov. Takšni programi pa morajo biti tudi zanesljivi in natančni, saj ima le ponovljiva in natančna znanost uporabno vrednost.

Znanstvena programska oprema se od tiste bolj poljudne razlikuje po tem, da je v glavnem ne pišejo programerji, ampak jo pišejo predvsem raziskovalci sami, ki za svoje raziskovalno vprašanje ne najdejo že obstoječega programa, zato se odločijo, da ga bodo napisali sami. Raziskovalci iz različnih področij akademske sfere (torej npr. fiziki, kemiki, biologi in drugi) pa v veliki večini nimajo dovolj izkušenj z delom na področju razvoja programske opreme. Zato je poudarek predvsem na tem, da programska oprema deluje in ne toliko na tem kako je napisana – pogosto ni dodatno komentirana, je slabo (ali pa sploh ni) dokumentirana, na voljo ni primerov uporabe, postopek prevajanja je zahteven in neznačilen. Večinoma niti ni pričakovana širša uporaba zunaj neke manjše raziskovalne skupine, kjer se navodila o prevajanju in uporabi prenašajo s pomočjo ljudskega izročila, pa se potem (bolj kot ne »po nesreči«) znajde v uporabi širšega kroga ljudi.

Velik problem je tudi t.i. pekel odvisnosti (ang. dependency hell), s čimer opišemo težave, ki se pojavijo, ko programska oprema sloni na številnih knjižnicah in paketih, le-ti pa imajo številne druge odvisnosti. Hkrati lahko določeni deli programske opreme zahtevajo različne verzije iste knjižnice ali pa za prevajanje potrebujemo določene različice prevajalnika, ki med sabo niso združljive. Tako se hitro soočimo s prepleteno mrežo zahtev, ki jo težko naslovimo, hkrati pa to vpliva tudi na ponovljivost znanstvenih računskih poskusov. Namreč, če natančne različice vseh odvisnosti niso dokumentirane in na voljo, je skoraj nemogoče ponoviti izračun naknadno ali na drugem sistemu.

Za obvladovanje številnih težav ob prevajanju programske opreme na superračunalnikih, se pojavlja nekaj rešitev. Prva so upravljavci paketov (ang. package managers), kot so EasyBuild, Spack, Conda, Nix ali GuixHPC. Na slovenskem superračunalniku Vega (kot tudi na večini ostalih superračunalnikov pod okriljem EuroHPC JU) je v uporabi EasyBuild, ki je namenjen enostavnejši namestitvi programske opreme iz izvorne kode, s pomočjo EasyBuild receptov, ki na poenoten način popišejo podporno programsko opremo in tudi okoljske spremenljivke programske opreme, kar bistveno zmanjša zahtevnost prevajanja in omogoča ponovljivost na različnih gručah. V konzorciju SLING imamo že od leta 2015 postavljen centralni CVMFS (Cern VM File System) strežnik z EasyBuild moduli, ki so na voljo na vseh slovenskih gručah, ki so del SLING-a.

Hkrati se v zadnjih letih financirajo evropski projekti, katerih namen je pomoč pri opisani problematiki. Tako se je leta 2024 začel EPICURE (ang. European Partnership for Innovative Computing in Research and Education), v okviru katerega strokovnjaki iz različnih evropskih gruč nudijo podporo pri pripravi in optimizaciji programske opreme za evropske uporabnike. Obstajajo pa tudi številni centri odličnosti (ang. Centers of Excellence), ki so namenjeni podpori in skaliranju znanstvene programske opreme iz različnih

domenskih področij, npr. biologije (BioExcel-3), materialov (MaX), večskalnih simulacij (MultiXscale), astrofizike (SPACE), itn.

Potrebno pa je omeniti še vsebnike (ang. containers), s pomočjo katerih si lahko uporabniki zagotovijo izolirano okolje, v katerem lahko poganjajo svojo programsko opremo. Hkrati so prenosljivi in omogočajo ponovljivost izvedenih računskih poskusov v istem okolju. Še ena prednost vsebnikov pa je tudi, da njihove definicije lahko prevzamejo vlogo dokumentacije.

