



# Če bomo v Evropi zaostali, bomo prišli pod oblast drugih

**Aleš Bošnjak, direktor Izuma** *Moč vege je enaka moči 40.000 osebnih računalnikov, moč novega superračunalnika bo 16-krat večja – Namenjen bo razvoju umetne inteligence*

Slovenija bo v Mariboru postavila še drugi superračunalnik, potem ko je leta 2021 tam zaživel HPC Vega, saj je bila država uspešna na evropskem razpisu skupnega podjetja EuroHPC za pridobivanje sredstev za superračunalnik (HPC) s tovarno umetne inteligence (UI). Projekt je vreden do 150 milijonov evrov. Bistvena značilnost novega superračunalnika je, da bo v celoti prilagojen storitvam in rešitvam UI. Po načrtih bo začel delovati v začetku leta 2027, do takrat bo deloval Vega. O projektu smo govorili z dr. Alešem Bošnjakom, direktorjem Izuma.

**SAŠA SENICA**

Prijava je bila zelo dobro točkovana, država je od Evrope pričakovala do 75 milijonov evrov, po podatkih iz Bruslja pa naj bi prejela 67,6 milijona evrov. Ste zadovoljni?

Zdaj sledijo pogajanja, tako da se lahko zgodijo prerazporeditve glede na skupni evropski ekosistem, zato o točnih številkah v pogodbi še ne morem govoriti. Projekt je vreden do 150 milijonov evrov, polovico naj bi dobili iz Evrope, drugo polovico pa bo financirala država. Ta znesek pokriva superračunalnik (strojna oprema, stroški delovanja, zavarovanje in zapolnitve dodatnih strokovnjakov) skupaj s tovarno UI in vložek v superračunalnik v Italiji, vključeno so morebitne dodatne prilagoditve v podatkovnem centru glede na dobavljeno tehnologijo.

**Ali superračunalniki, ker gre za temeljno orodje za razvoj UI, že spadajo pod kritično infrastrukturo?**

To ne. Po novem znanstveno-raziskovalnem zakonu Izum gosti kritično infrastrukturo, ki jo verjetno vsi poznajo, to sta Cobiss in raziskovalni informacijski sistem Sicris. Če bi sistem Cobiss padel, bi bile vse knjižnice ohromljene. Po novem zakonu imamo zato za ta sistem lahko definirano stalno pripravljenost, saj se moramo

odzvati v kateremkoli trenutku, če bi sistem javil napako.

Za superračunalnik pa to za zdaj ni potrebno. V primeru izpada recimo nekaj računalnikov oziroma vozlišč znotraj superračunalnika se lahko prekinjeni izračuni oziroma računski posli ponovijo. To sicer poveča strošek izračunov in podaljša čas izvedbe, vendar je vseeno to mogoče izvesti. Superračunalnik je treba razumeti kot infrastrukturo skupaj s podatkovnim centrom, podatki in drugo opremo, ki je namenjena temu, da raziskovalci naložijo svoje podatke in programsko opremo, s katero podatke obdelujejo, simulirajo nove procese, optimizirajo obstoječe ali izvajajo stranske izračune. Če pri tem ne gre za testiranje zmogljivosti ali razvojne aktivnosti, temu pravimo produkcijsko računanje, katerega končni cilj je hiter in energijsko najmanj potraten izračun. Ko raziskovalci dobijo rezultat, denimo v obliki modela UI, ga prenesejo in namestijo na neki drug strežnik, kjer bo živel in deloval. Superračunalniki so v primeru UI za zdaj namenjeni razvoju, testiranju in učenju modelov.

**Vendar pa gre za izjemno pomembno opremo za razvoj v celotni Evropi, ki zdaj pospešeno vlaga sredstva, da dohiti ZDA in Kitajsko.**

Seveda gre za zelo pomembno opremo, saj se UI na vseh področjih razvija z veliko hitrostjo, in Evropska unija se je tu upravičeno zbalala, da tega razvoja ne bomo dohajali, če ne bomo nemudoma skočili na vlak in v področje vložili ne le sredstva, ampak tudi ogromno znanja evropskih razvojno-raziskovalnih sil. Pri tem razvoju pa morajo sodelovati tudi podjetja.

Novi HPC, za razliko od starega vege, ne bo namenjen izključno raziskovalcem. Seveda bodo ti skrbeli za razvoj, a pomembno je predvsem to, da bodo rešitve prevzela, uporabljala in tudi prodajala podjetja. Pri tem bo zelo pomemb-

no izobraževanje, da bodo uporabniki to znali koristiti. Ustvariti bomo morali most med znanostjo, posredniki in končnimi uporabniki novih modelov. Prav zato je evropska komisija razpis prilagodila tako, da so bile v ospredju tiste vsebine, ki jih gospodarstvo in družba potrebuje v realnem sektorju. Tokrat ne bo šlo toliko za temeljne raziskave, ampak za zelo aplikativne.

**Evropa je precej zadaj pri razvoju UI. Kako vi vidite ta razvoj, bo Evropa dohitela omenjeni vesilici?**

Res je, ampak treba je ločiti različne tipe UI. Kot drugo, vsak model UI je lahko tudi krajevno odvisen. Evropi, kot je dejal namestnik direktorja EuroHPC JU Thomas Skordas, ne pomaga ameriški ali katarski model UI za kmetijstvo, saj je to področje v teh dveh državah popolnoma drugačno kot v Evropi. Podatki so drugačni, metodologija je drugačna, okoljske spremembe so drugačne, prav tako padavinski sistem, zato mora biti model za kmetijstvo evropski. Superračunalniški centri v Evropi ne bodo izolirani, ampak povezani. Na naslednjih sestankih se bomo dogovorili o oblikovanju delovnih skupin, zato se ne bomo podvajali.

**Če narediva korak nazaj in se najprej posvetiva superračunalniku Vega: kaj je ta prinesel slovenski znanosti, družbi?**

HPC Vega nam je zelo pomagal pri novem razpisu, saj smo si z njim in njegovim upravljanjem pridobili zaupanje evropske komisije. Našim znanstvenikom je Vega omogočil neposreden dostop do računske moči, ki je prej niso imeli. Mesečno se na HPC Vega izvaja med petdeset in sto projektov, ki so različno veliki, vsaj še enkrat toliko pa je aktivnih uporabnikov mesečno. HPC Vega je do zdaj uporabilo prek tisoč naših in tujih raziskovalcev, kljub starosti je še vedno zelo uporaben. Vseskozi je maksimalno zaseden, pri tem naj pohvalim naše zelo dobro sodelovanje z univerzama v Ljubljani in Mariboru in IJS, ki ima zelo

dobre izkušnje s sodelovanjem na podobnem superračunalniku v Cernu v Švici, ki so nam pomagale pri vzpostavljanju vege. Računska moč vege je razdeljena na 65 odstotkov za slovensko raziskovalno sfero in 35 odstotkov za evropsko.

**Bistvena razlika med vejo in novim HPC bo v sestavi enot. Vejo večinoma sestavljajo centralne procesne enote (CPU), le za tretjino je grafičnih procesnih enot (GPU). Kateri procesorski deli računalnika so namenjeni kateri nalogi ali za kaj se uporablja CPU in za kaj GPU?**

CPU je v resnici aritmetični procesor in lahko zelo hitro izvaja zelo veliko količino aritmetičnih operacij. Zato so namenjeni izračunom pri kakšnih temeljnih modelih, kjer imamo kompleksne enačbe, ki jih lahko razbijemo na enostavne aritmetične operacije. GPU pa je grafični procesor in je namenjen za računanje v grafičnem oziroma v vektorskem prostoru, kar je pomembno pri UI. Na primer: slika, ki jo prepoznavamo, je opisana z vektorji, navadne aritmetične operacije so praktično neuporabne. Enako velja za velike jezikovne modele, kjer se računa oddaljenost besede od besede in tako naprej, ali pa za modele, kjer napovedujemo na osnovi različnih geostrateških ali meteoroloških podatkov, pri čemer imamo opravka z zelo veliko količino podatkov.

Pri večini modelov UI delujejo grafični procesorji bistveno bolj učinkovito. Mi imamo na veji eno tretjino moči v grafičnih procesorjih, vse ostalo v aritmetičnih. Namenjen je bolj raziskovalcem kot industriji, čeprav ga nekatera podjetja koristijo pri svojih izračunih za plačilo. Temu pravimo komercialni dostop in je dovoljen do 20 odstotkov kapacitet superračunalnika.

**Novi bo imel samo GPU?**

Imel bo tudi dovolj zmogljivo particijo samo s procesorji CPU. Kljub temu bo največ računske moči na vozliščih z GPU, torej osnovni računalniki v gručah bodo imeli zmogljive grafične pospeševalnike. Treba je vedeti, da se ta tehnologija spreminja iz dneva v dan. Ko smo začeli s projektom Vega, smo najprej pripravili zasnovano za zračno hlajenje procesorjev, vendar so ti v kratkem času postali toliko zmogljivejši, da smo se osredotočili na vodno hlajenje.

Dinamika je zelo velika, zato je naš projekt zastavljen zelo modro

v dveh fazah. Prva vključuje naročilo, dobavo in vzpostavitev strojne opreme ter delovanje tovarne umetne inteligence. Druga faza, predvidena dve leti po začetku delovanja, pa bo obsegala nadgradnjo strojne opreme z najnovejšimi grafičnimi procesnimi enotami in drugo tehnologijo za dolgoročno konkurenčnost superračunalnika. S tem smo si tudi podaljšali garancijo.

**Razvoj je res zelo hiter. Kako drugačni so bili denimo ti računalniki pred dvajsetimi leti?**

Popolnoma drugačni. Na primer: vsi računalniški sistemi, ki so pomagali Apollu, da je pristal na Luni, so le majhen del tega, kar imamo danes v telefonu. Tudi moderni superračunalniki so neprimerno zmogljivejši, še pred leti smo govorili o nekaj jedrih, danes jih je že 200 in več, prav tako vsebujejo neprimerno več vozlišč z več procesorji na enem vozlišču, obenem imajo več pomnilnika in vsebujejo grafične pospeševalnike, ki jih pred leti v klasičnih superračunalnikih niti ni bilo. Zelo pomembna je tudi hitrost omrežja, ki povezuje vozlišča, ta je narasla z ena Gbit/s vse do 800 Gbit/s. Tudi kapacitete in zmogljivosti diskovnih sistemov za shranjevanje so danes neprimerno večje kot pred leti.

**Vega ima zmogljivost 6,9 petaflopov na sekundo, prihodnji HPC pa bo imel zmogljivost 110 petaflopov – petaflop je število operacij s plavajočo vejico na sekundo. Kako bi najlažje razložili, kaj to pomeni?**

Vega, ker ga lahko primerjamo z aritmetičnimi procesorji osebnih računalnikov, ki imajo običajno štirijedrno procesorje, nekateri tudi več, ima moč približno 40.000 takih osebnih računalnikov. No, novi bo imel 16-krat večjo moč.

**Torej, kot 640.000 osebnih računalnikov?**

A ker gre za GPU, gre pravzaprav še za več računske moči.

FLOPS (floating-point operations per second) je sicer enota, s katero se meri zmogljivost procesorjev, računalnikov in superračunalnikov. Predpona peta- pomeni 10<sup>15</sup> takih operacij. Vendar naj opozorim, da že nekaj časa tako izmerjena zmogljivost ni pravi pokazatelj računske moči superračunalnikov, ki so prilagojeni za UI. Prihodnji HPC bo imel okoli 10 EFLOPS (eksa je 10<sup>18</sup>) v mešanem načinu FP8 (floating-point 8-bit) na delu za pospeševanje izračunov

v projektih UI. Vega ima 240 grafičnih pospeševalnikov, novi HPC bo imel 2000 novejših grafičnih pospeševalnikov predvsem za učenje modelov, obenem pa tudi vsaj 100 strežnikov z grafičnimi podsistemi, ki bodo omogočali zaganjanje, uporabo in preskušanje UI v praksi.

Novi superračunalnik bo vseboval zmogljive diskovne sisteme s kapaciteto 110 petabajtov (PB – en petabajt je enako kot milijon gigabajtov), kar je petkrat več kot pri HPC Vega. Povezljivost v Slovenijo in Evropo pa bo več kot 1 Tbit/s. Obenem bo v istem podatkovnem centru v Mariboru v prvem nadstropju tudi oprema Arnesa za trajno hrambo podatkov z začetno kapaciteto 100 PB ter strežniki nacionalne strukture v oblaku za raziskovalne organizacije in javni sektor.

**Enote GPU bodo iz podjetja Nvidia?**

O tem še ni odločitve. Vendar je na prejšnjem razpisu najboljšo rešitve ponujala Nvidia. Menim, da bo tudi tokrat tako. Nvidia je najbolj prodorna, ima velik tržni delež in veliko podporne programske opreme, grafični pospeševalniki so tudi najbolj razširjeni in uporabljani. Velika slabost je, da so lahko zato cene višje, pride lahko tudi do težav z dobavo opreme, saj je povpraševanje večje od proizvodnje. Pri tem pa bi lahko vplivala tudi geopolitična situacija, saj predsednik Trump iz dneva v dan sprejema nove odločitve.

**Celotno infrastrukturo pa bo verjetno sestavljalo več podjetij?**

Da, odvisno od tega, kdo se bo prijavil na javni razpis. Vsekakor bo šlo za priznanega dobavitelja in vzdrževalca te napredne opreme. Ponuditi bo moral vse, od hlajenja, vzdrževanja strojne opreme v garancijskem obdobju do drugih storitev, ki jih bomo zahtevali. Povsem normalno je, da bo imel podizvajalce, ki bodo lahko ali domača ali tuja podjetja. Tudi novi superračunalnik se bo hladil s tekočino, del opreme, kot so diskovni sistemi, pa se bo hladil klasično z vpihovanjem hladnega zraka.

**Novi HPC bo postavljen ob elektrarni Mariborski otok, gradbeno dovoljenje je že izdano. K odločitvi je predvidevam prispevalo predvsem to, da je tam elektrarna. Kako ste se soočili z izzivom, da bi HPC deloval čim bolj varčno oziroma da bo projekt trajnosten?**

Da, bližina vira okoljsko čiste energije je bil pomemben faktor pri izbiri lokacije, da ne bo dodatno obremenjeno distribucijsko omrežje. HE Mariborski otok ima priključno moč 60 MW, novi podatkovni center Arnesa (v katerem bo tudi HPC) bo priključen na največjo moč okoli 10 MW, v izračunih smo predvidevali, da za superračunalnik ne bo potrebno več kot nekje 4,5 MW moči; je pa to odvisno od tehnologije in uporabe superračunalnika.

Procesorji v superračunalnikih so izkoriščeni med 80- in 90-odstotno, medtem ko so procesorji v osebnih računalnikih izkoriščeni le med 10 in 15 odstotki, prav zato se v petih letih delovanja iztrošijo. Porabljena energija za delovanje in hlajenje se večinoma spremeni v toplotno energijo, zato Dravske elektrarne in Energetika Maribor razmišljajo o vzpostavitvi omrežja za distribucijo toplotne energije za ogrevanje dela mesta. Ne vem, kako daleč so s tem projektom. Nekaj tveganja sicer je, če po sedmih letih ne bi več dobili novega superračunalnika, bi ta infrastruktura ostala neuporabljena.

Gradbeno dovoljenje je že izdano in menim, da bo gradnja stekla zelo kmalu, saj mora biti stavba pripravljena na prihod novega HPC.

#### **Koliko znaša račun za elektriko za delovanje vege in koliko bo znašal računa za novi HPC?**

Mesečni računi so od začetka delovanje vege znašali od 50.000 do 120.000 evrov, odvisno od cene, ki jo pridobimo na javnem razpisu, in od delovanja vege. Grobe izračune za električno energijo smo morali prikazati tudi v prijavi, saj bo EuroHPC JU pokrival stroške energije za tisti del rabe, ki bo namenjen podjetjem in organizacijam iz Evrope, med njimi pa so lahko tudi naša. V sedmih letih delovanja se predvideva okoli 30 milijonov evrov stroškov za elektriko.

#### **Kar nekaj strokovnjakov ste morali izobraziti za vzdrževanje vege in tudi za podporo raziskovalcem. Koliko ljudi boste potrebovali, da bo novi HPC delal brez težav in da bo podpora dobra? Kako resen je pravzaprav kadrovski problem?**

Pri nas ne potrebujemo le strokovnjakov IT, imeli bomo ogromno projektov, ki jih bo treba voditi na primeren način, tako bomo potrebovali zelo dobre administratorje projektov, pa pravne strokovnjake, ekonomiste ... Za samo podporo uporabnikom in za vzdrževanje novega superračunalnika sicer

predvidevamo med deset in petnajst novih zaposlitev na Izumu, IJS in Arnesu. Pri projektu UI pa imamo predvidene še triletno zaposlitve posameznih raziskovalcev.

Strokovnjake IT in programerje je zelo težko dobiti, saj jih primanjkuje v celotni državi, določenih strokovnjakov za UI pa sploh nimamo. Maribor je blizu Avstrije in strokovnjaki odhajajo, ker imajo v sosednji državi višje plače. Mi zelo veliko investiramo v študentski kader, ki mu omogočamo prakso, in pogosto ti študenti pri nas tudi ostanejo, če je skupni interes. Tu zelo dobro sodelujemo z mariborsko univerzo. Nenehno se moramo sicer truditi, da zagotavljamo dobre delovne razmere in solidno plačo, kolikor je to v javnem sektorju sploh mogoče. Vsekakor so v zasebnem sektorju te mnogo mnogo višje.

#### **Del celotnega projekta je tudi tovarna umetne inteligence, naša je imenovana SLAIF. Ime nekoliko zmede, saj ne gre za tovarno kot tako, bolj za konzorcij največjih strokovnjakov na področju UI v Sloveniji.**

Ne samo strokovnjakov, ampak tudi podjetij. Menim, da bi tovarno UI še najbolj opisali kot valilnico modelov ali storitev UI, ki so namenjeni uporabnikom. Torej je s pojmom »tovarna« mišljena vsebina, ki bo nastajala na računalnikih. Raziskovalci in podjetja bodo prijavljali projekte, pri tem je mišljeno, da bodo podjetja projekte tudi ustrezno sofinancirala. Slovensko nacionalno superračunalniško omrežje (SLING), ki je konzorcij partnerjev (vodi ga Arnes) za razvoj superračunalništva, omrežnega oz. razpršenega računalništva ter visokozmogljivih računalniških sistemov za uporabo pri raziskavah v znanosti, izobraževanju in gospodarstvu ter javni upravi, bo odločalo, kateri projekti so prioritetni, svojo komisijo ima tudi EuroHPC JU, ki bo odgovorna za tisti del projektov, ki jih bo financirala EU.

Razdelitev računalniške moči bo po novem 50 – 50, ker bo tokrat HPC namenjen predvsem podjetjem, ki pa bodo tesno sodelovala z raziskovalci. Ti bodo imeli dostop do opreme še naprej pod enakimi pogoji kot doslej. Raziskovalnih skupin v Sloveniji je seveda kar nekaj.

#### **Vi sicer veliko uporabljate UI?**

Ja, kar. Na različnih ravneh. Programerjem denimo chatgpt vsekakor

pomaga, da hitreje spišejo kodo. Pogovarjamo se tudi, da bi s pomočjo UI podprli procese katalogizacij, ki jih potrebujemo v knjižnicah.

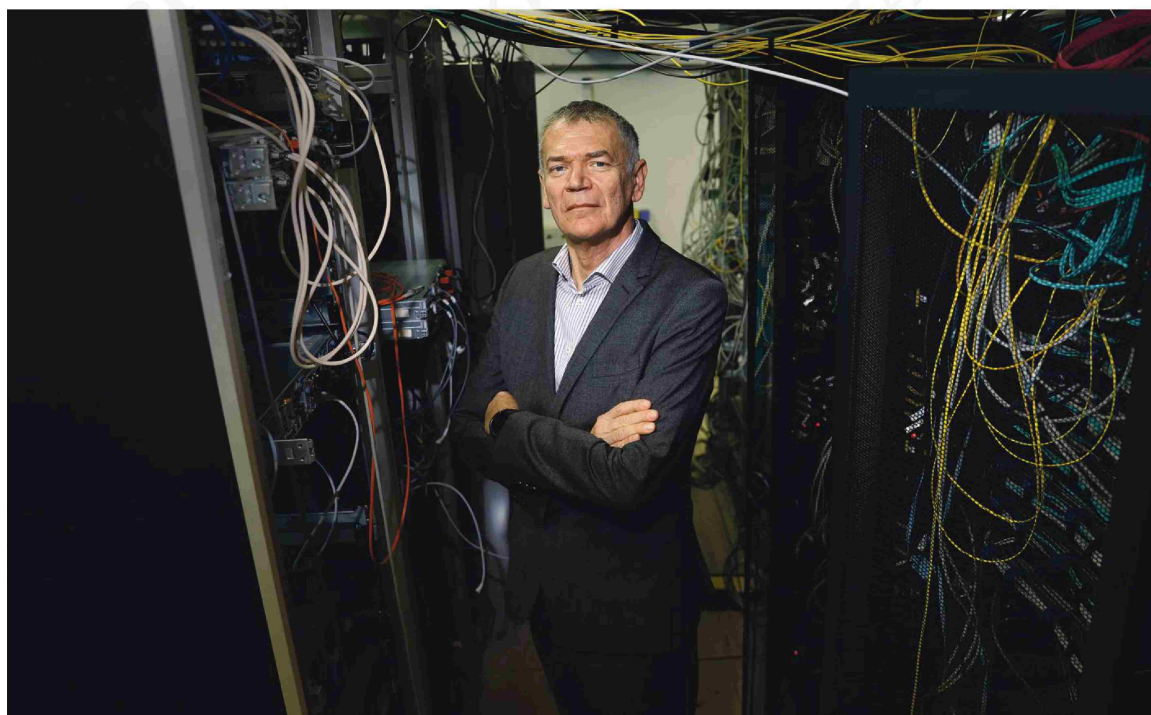
Vsekakor gre za eksplozijo razvoja in uporabe, in če bomo v Evropi zaostali, bomo vsebinsko in znanstveno prišli pod oblast drugih, ki bodo to imeli. Če hoče Evropa ostati samostojna, mora sama obvladovati te medele in jih mora biti sposobna razvijati, ker jih Američani za nas ne bodo, tam je drugo okolje in so seveda tudi drugi interesi.

**Če je dobil prejšnji HPC ime po Juriju Vegi, bo ta dobil ime po ...?**  
Za zdaj ga imenujemo kar Vega 2, morda bo kasneje dobil drugačno ime.

## SODELUJOČE USTANOVE

Novi superračunalnik bo Institut za informacijske znanosti (Izum) v sodelovanju z Institutom Jožef Stefan (IJS) in Arnesom namestil v novem podatkovnem centru, ki ga bo Arnes zgradil na območju Dravskih elektrarn v Mariboru. V projektu slovenske tovarne UI sodelujejo kot partnerji IJS, Univerza v Ljubljani, Univerza v Mariboru, Univerza v Novi Gorici, Izum, Arnes, Univerza na Primorskem, Fakulteta za informacijske študije Novo mesto, Gospodarska zbornica Slovenije in Tehnološki park Ljubljana. Izum s podporo ministrstva za digitalno preobrazbo je kot superračunalniški center tudi uradni koordinator prijave, IJS pa koordinira projekt tovarne UI. V njej bodo razvijali predvsem rešitve za zeleni prehod, zdravje in biotehnologijo (diagnostika, personalizirana medicina, odkrivanje zdravil, orodja za odločanje), digitalno družbo (jezikovne tehnologije, generativna UI, preobrazba javnega sektorja, izobraževanje) in znanost (znanstveno modeliranje in avtomatizacija, razvoj materialov, okoljska znanost, digitalna humanistika).

*Novi HPC, za razliko od starega vege, ne bo namenjen izključno raziskovalcem. Seveda bodo ti skrbeli za razvoj, a pomembno je predvsem to, da bodo rešitve prevzela, uporabljala in tudi prodajala podjetja.*



Novi superračunalnik bo vseboval zmogljive diskovne sisteme s kapaciteto 110 petabajtov, pravi Aleš Bošnjak, direktor Izuma. FOTO LEON VIDIC