# Zdůvodnění užití obchodního označení výrobku (viz také odst. 4.3 zadávací dokumentace)

Komplementární systémy mají za cíl umožnit uživatelské komunitě přístup k technologiím a architekturám, které jsou v ČR dosud nezvyklé nebo nedostupné, současně se však jedná o technologie perspektivní nebo široce akceptované zejména v Asii a ve Spojených státech. Rozšíření dostupného prostoru výpočetních technologií je důležité pro robustnost a konkurenceschopnost výzkumného prostředí v ČR. Komplementární systémy poskytnou vědeckým týmům příležitost vyniknout v nových doménách výzkumu.

Výpočetním technologiím v ČR a Evropě dominují architektury x86\_64 a Nvidia GPGPU. Postupné investice umožňují zahrnout do Projektu i menší komplementární clustery, pomocí kterých bude uživatelům nabídnut přístup k dalším výpočetním architekturám, které se v současné době aktivně rozvíjejí v oblasti HPC.

V komplementárních systémech I. bylo portfolio nabízených systémů rozšířeno o ARM 8.2 systém založený na procesoru A64FX, který je srdcem počítače Fugaku, servery s AMD CPU, GP-GPU a Xilinx FPGA a servery s Intel CPU a Intel FPGA.

Komplementární systémy I, však stále nepokrývají všechny důležité technologie, ke kterým nemají naši uživatelé běžný přístup. Proto je nadále rozšiřujeme o následující technologie v rámci veřejné zakázky Komplementární systémy II: (i) kombinace ARM procesorů s GPU akcelerátory a současně s programovatelnými síťovými akcelerátory (DPU), (ii) nové typy procesorů, které díky chipletové technologii mají řádově větší cache, (iii) výkonné procesory založené na Power architektuře, zde konkrétně Power10 a (iv) Virtuální GPU akcelerované pracovní stanice.

V souladu s cíli projektu je smyslem v rámci tohoto návrhu uživatelům e-Infra CZ nabídnout mimo jiné tři nové technologie, u kterých používáme obchodní názvy.

První z nich je technologie procesorů pod obchodním názvem Power10. Procesory založené na této architektuře jsou mimo jiné používány v USA, kde jsou na této technologii vybudovány počítače Summit či Sierra (Summit byl č.1 do roku 2019 a současně je na 4. pozici, Sierra o stupeň horší). Jedná se tedy jak historicky, tak do budoucna o důležitou HPC technologii a je nutné nabídnou uživatelům vývojové prostředí. Procesory této architektury jsou dostupné pouze od výrobce IBM a není dostupná alternativa.

Druhou z nich jsou nové typy procesorů, které díky chipletové technologii mají řádově větší cache. S nástupem chipletové technologie začal výrobce procesorů AMD jako první nabízet procesor s obrovskou L3 cache. Jedná se o parametr, který zásadně ovlivní optimalizace kódu pro výpočetně náročné úlohy. Tato partice umožní uživatelům připravit jejich kódy na tento nový typ procesorů. V současné době je jediný výrobce těchto procesorů AMD a jeho model Milan-X.

Poslední technologií je virtuální GPU akcelerované pracovní stanice. Tyto jsou primárně určeny uživatelům pro efektivní „pre-processing“ a „post-processing“ úloh nebo modelů pro rozsáhlé simulace nebo výpočty. Cílem je dát uživatelům prostředky pro velmi výkonné virtuální prostředí se schopností přenosu obrazu s minimální latencí a maximálním výkonem – tedy snahou je poskytnout maximální uživatelský komfort pro pohodlnou práci se vzdálenou pracovní plochou, s hardwarovou akcelerací 3D aplikací. Konkrétně cílíme na softwarové řešení NVIDIA RTX Virtual Workstation (vWS). Díky tomu, že GPU hardware, jeho specializované uzavřené ovladače a uzavřená nadstavba, která poskytuje požadovanou službu jsou od stejného výrobce, dosahuje toto řešení nejlepšího výkonu. Podobné řešení se dá sestavit i pomocí open-source řešení (VirtualGL a VNC server a client), které ale nedosahují stability a výkonu požadovaného řešení a neumožnuje sdílení zdrojů grafických karet pro více uživatelů. Pro národní superpočítačové centrum je mimořádně důležitá schopnost bezpečného sdílení výpočetních zdrojů pro celou akademickou komunitu České republiky což je unikátní schopnost požadovaného řešení.

Obchodní označení konkrétních výrobků bylo užito s cílem zajistit naplnění cílů projektu, z důvodů technických vazeb a technologických výlučností je z výše uvedených důvodu potřebné cílit na konkrétní produkty.

## **Partice 1:** Kombinace ARM procesorů s GPU akcelerátory a současně s programovatelnými síťovými akcelerátory (DPU)

Separátně ARM procesory a NVIDIA GPU akcelerátory jsou klíčové technologie a základní stavební kameny největších počítačů světa (Japonsko – Fugaku /ARM/, USA – Summit /GPU/, Německo – Juwels Booster /GPU/). V roce 2023 plánuje Švýcarské centrum CSCS rozšířit superpočítač ALPS[[1]](#footnote-2) o výpočetní uzly obsahující procesory GRACE úzce propojující ARM procesory s GPU akcelerátory. Tato partice potom umožní portování kódů na tento typ počítačů.

Současně tato partice nabízí uživatelům přístup k HW umožňující vývoj aplikací podporující tzv. „in network computing“, tedy zpracování dat pomocí síťových prvků. Zde se konkrétně jedná o chytré síťové karty (DPU)[[2]](#footnote-3).

Jelikož se jedná o několik různých technologií, požadujeme řešení, kde klíčové elementy (GPU, DPU) jsou od jednoho výrobce a ARM CPU je tímto výrobcem plně podporován[[3]](#footnote-4).

## **Partice 2:** Nové typy procesorů, které díky chipletové technologii mají řádově větší cache.

S nástupem chipletové technologie začal výrobce procesorů AMD jako první nabízet procesor s obrovskou L3 cache. Jedná se o parametr, který zásadně ovlivní optimalizace kódu pro výpočetně náročné úlohy. Tato partice umožní uživatelům připravit jejich kódy na tento nový typ procesorů.

## **Partice 3:** Výkonné procesory založené na Power architektuře, zde konkrétně Power10.

Procesory založené na architektuře Power jsou mimo jiné používány v USA, kde jsou této technologii vybudovány počítače Summit či Sierra (Summit byl č.1 do roku 2019 a současně je na 4. pozici, Sierra o stupeň horší). Jedná se tedy jak historicky tak do budoucna o důležitou HPC technologii a je nutné nabídnou uživatelům vývojové prostředí.

## **Partice 4:** Virtuální GPU akcelerované pracovní stanice.

Jsou primárně určeny uživatelům pro efektivní „pre-processing“ a „post-processing“ úloh nebo modelů pro rozsáhlé simulace nebo výpočty. Cílem je dát uživatelům prostředky pro velmi výkonné virtuální prostředí se schopností přenosu obrazu s minimální latencí a maximálním výkonem – tedy snahou je poskytnout maximální uživatelský komfort pro pohodlnou práci se vzdálenou pracovní plochou, s hardwarovou akcelerací 3D aplikací.

Obchodní označení konkrétních výrobků bylo užito s cílem zajistit naplnění cílů projektu, z důvodů technických vazeb a technologických výlučností je z výše uvedených důvodu potřebné cílit na konkrétní produkty.

1. ALPS Supercomputer <https://www.cscs.ch/science/computer-science-hpc/2021/cscs-hewlett-packard-enterprise-and-nvidia-announce-worlds-most-powerful-ai-capable-supercomputer/> [↑](#footnote-ref-2)
2. Data Processing Units (DPU) <https://www.nvidia.com/en-us/networking/products/data-processing-unit/> [↑](#footnote-ref-3)
3. Arm HPC Developer Kit <https://developer.nvidia.com/arm-hpc-devkit> [↑](#footnote-ref-4)