

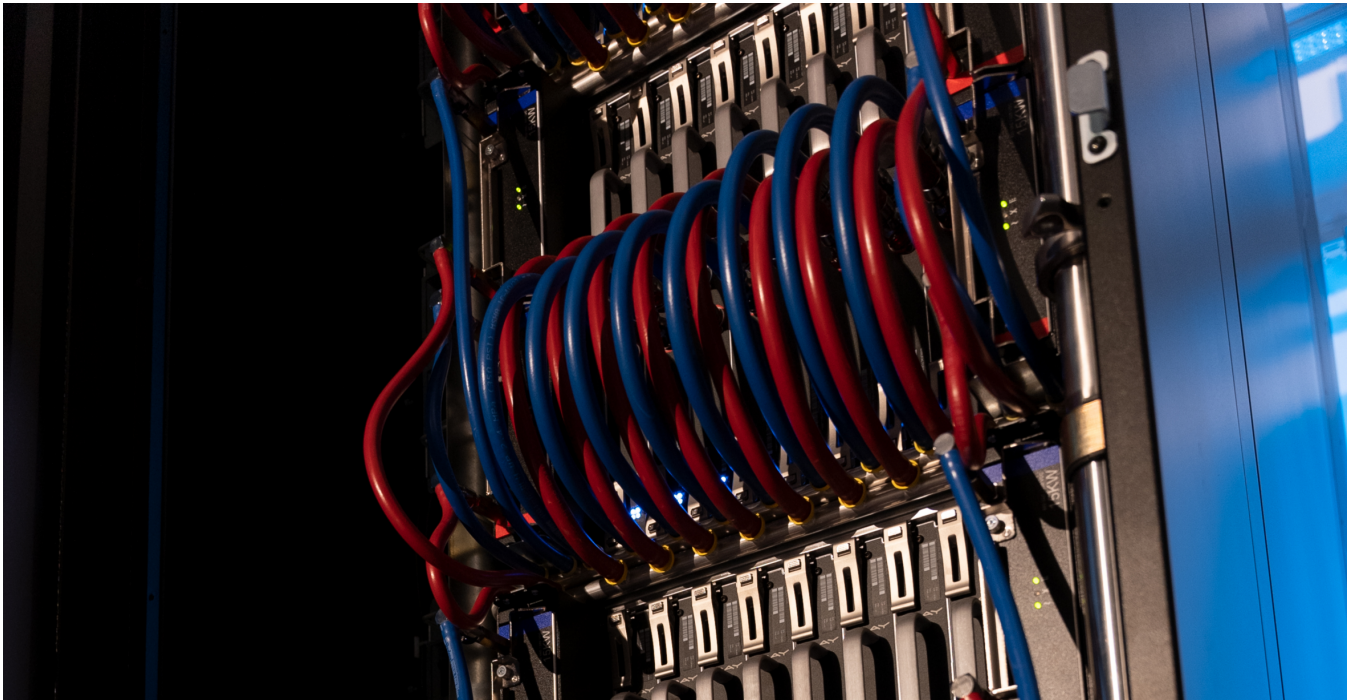
Arhitektura Supeka

- Uvod
- Procesori
- Poslužitelji
 - Pristupni poslužitelji
 - CPU radni poslužitelj
 - GPU radni poslužitelj
 - Poslužitelji s velikim memorijskim kapacitetom
- Međuveza
- Podatkovno spremište

Uvod

Glavnina procesorske snage Supeka temelji se na tehnologiji *HPE Cray EX*. Ova tehnologija temelji se na arhitekturi *Shasta* i uključuje niz inovativnih značajki koje su namijenjene za upotrebu u vrlo velikim i zahtjevnim računalnim aplikacijama, uključujući simulacije, modeliranje, istraživanje, genetiku i druge znanstvene i poslovne aplikacije. Jedna od ključnih značajki Supeka je njihov vrhunski interkonekcijski sustav, koji koristi *Cray Slingshot* mrežu. Ova mreža omogućuje brzu razmjenu podataka između čvorova, što je ključno za vrlo brzo izvođenje računalnih operacija. Osim toga, Supek ima i visoko učinkovit sustav hlađenja, koji omogućuje održavanje optimalne temperature unutar računalnog sustava, čime se osigurava vrhunska izvedba čak i u najzahtjevnijim uvjetima rada. Ova vrsta hlađenja omogućuje efikasno odstranjivanje topline vodom putem sistema cijevi i izmjenjivača topline tzv. *cooling distribution unit (CDU)*. Superračunalo Supek sastoji se od više komponentata koje tipično nalazimo na [računalnim klasterima](#):

- Pristup - Poslužitelji namijenjeni korisničkoj interakciji s cijelim klasterom
- Procesori - Superračunalni dio koji sadrži procesorske jezgre
- Podatkovno spremište - Centralni spremišni dio koji je dostupan cijeloj mreži
- Međuveza - Mreža visoke učinkovitosti koja povezuje sve dijelove



Ostranjivanje topline vodom putem sistema cijevi

Procesori

Svi poslužitelji na klasteru Supek sadrže jedan ili dva AMD EPYC 7763 CPU-a.

AMD EPYC 7763 je dio serije procesora [Epyc 7003 Milan](#) razvijen od kompanije AMD, izgrađen na arhitekturi Zen 3, koja omogućava bolju energetska učinkovitost i povećava ukupne performanse u odnosu na prethodne generacije. Procesor je izgrađen na 7-nanometarskoj proizvodnoj tehnologiji, što ga čini vrlo učinkovitim u potrošnji energije i toplinskom upravljanju. AMD EPYC 7763 procesor također ima mogućnosti dinamičke prilagodbe snage koja prilagođava radnu snagu procesora kako bi se optimizirala energetska učinkovitost. To je korisno za okruženja poslužitelja koja trebaju optimalnu snagu obrade, ali i žele smanjiti potrošnju energije.

Procesor se sastoji od 64 jezgre, a jezgre su u potpunosti kompatibilne sa X86-64 arhitekturom i podržavaju AVX2 256-bit vektorske instrukcije sa maksimalnom propusnosti od 16 "double precision" FLOP-a/takt (AVX2 FMA operacije) ili ukupno 2,5 teraFLOPS-a pri osnovnom radnom taktu od 2,45 GHz po jednom procesoru.

Specifikacije procesora AMD EPYC 7763 su sljedeće:

- Broj jezgara: 64
- Broj dretvi: 128
- Osnovni takt: 2,45 GHz
- Maksimalni takt: 3,5 GHz
- Cache memorija: L3 - 256 MB, L2 - 512 kB, L1 - 32 kB
- TDP: 280 W
- Podrška DDR4 memorijske module do 3200 MHz
- Podrška do osam kanala DDR4 memorije
- PCIe verzija: 4.0

Na računalnom klasteru Supek nalazi se ukupno 80 NVIDIA A100 40GB GPU-a u SXM izvedbi na radnim poslužiteljima i jedan GPU iste serije u PCI izvedbi na pristupnom poslužitelju.

NVIDIA A100 40GB je grafička kartica koja je posebno dizajnirana za izvođenje zahtjevnih računalnih operacija, kao što su znanstveno računanje, strojno učenje i visoko učinkovito računanje. Zahvaljujući svojoj arhitekturi [Ampere](#), NVIDIA A100 40GB omogućava poboljšanu obradu podataka i performanse u usporedbi s prethodnim NVIDIA grafičkim karticama. Njena specifikacija uključuje:

- Arhitektura: Ampere
- Procesor: NVIDIA A100 Tensor Core GPU
- Broj CUDA jezgara: 6.912; razne veličine instanci do 7 MIG-a @ 5GB
- Broj Tensor jezgara: 432
- Memorija: 40 GB
- Tip memorije: [HBM2](#)
- Sabirnica: 5120 bit
- Propusnost: 1555 GB/s
- TDP: 500W (2000W)

Procesor *NVIDIA A100 Tensor Core GPU* sastoji se od 6.912 CUDA jezgara i 432 *Tensor* jezgara. Razlika između *CUDA* i *Tensor* jezgri može se vidjeti u njihovoj primarnoj funkciji. CUDA jezgre se koriste za paralelno izvođenje širokog raspona algoritama za obradu slika, znanstveno računanje i mnoge druge aplikacije koje se mogu paralelizirati. *Tensor* jezgre su posebne jezgre koje se koriste za obradu tenzora. Ove jezgre pomažu u brzom izvođenju složenih matematičkih operacija, što je ključno za izvođenje zahtjevnih operacija strojnog učenja.

Ukupni kapacitet memorije grafičke kartice je 40GB. Ova količina memorije omogućava brzo pohranjivanje velikih količina podataka koje se koriste u zahtjevnim računalnim aplikacijama. To znači da korisnici mogu obraditi velike količine podataka i smanjiti vrijeme potrebno za izvođenje računalnih operacija.

Poslužitelji

Specifična izvedba koja čini Supek je HPE Cray EX2500 i sastoji se od više poslužitelja:

Namjena	Broj	CPU	GPU	RAM (GB)
CPU pristupni poslužitelj	1	2 x AMD EPYC 7763	-	256
GPU pristupni poslužitelj	1	1 x AMD EPYC 7763	1 x NVIDIA A100 (PCI)	128
CPU radni poslužitelj	52	2 x AMD EPYC 7763	-	256
GPU radni poslužitelj	20	1 X AMD EPYC 7763	4 X NVIDIA A100 (SXM)	512
Poslužitelji velikog mem. kapac.	2	2 X AMD EPYC 7763	-	4096

Pristupni poslužitelji

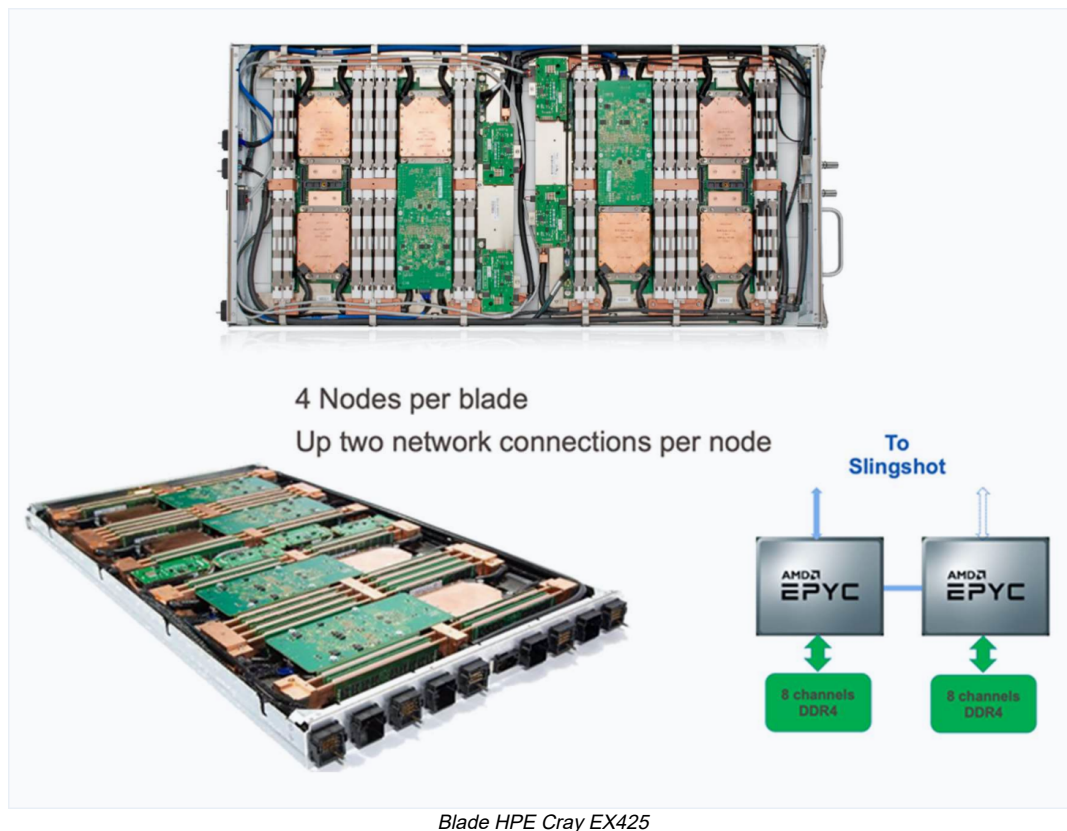
Na klasteru Supek su dostupna 2 pristupna poslužitelja/čvora.

Prvi pristupni čvor **HPE Proliant DL385 Gen10 Plus v2 (CPU pristupni poslužitelj)** ne sadrži grafički procesor ali sadrži **2 AMD EPYC 7763** procesora sa 64 jezgre (**128 jezgri ukupno**), matične ploče sa 8 memorijskih utora po procesoru maksimalne brzine prijenosa od 3200MT/s i ukupno 16 DDR4 memorijska modula od 16GB radne memorije (**256GB ukupno**). Poslužitelj je opremljen sa jednim lokalnim **NVMe SSD-om** kapaciteta **1,92 TB** i sa dodatnih **2 NVMe SSD-a** kapaciteta **7,68 TB** svaki.

Drugi pristupni čvor (**GPU pristupni poslužitelj**) je sustav **HPE Apollo 6500 Gen10 Plus** sa čvorom **HPE Proliant XL645d Gen10 Plus**. Poslužitelj je opremljen sa jednim grafičkim procesorom **NVIDIA A100 u izvedbi PCI** i jednim procesorom **AMD EPYC 7763** sa **64** procesorske jezgre. Matična ploča poslužitelja sadrži 8 memorijskih utora po procesoru i podržava brzinu prijenosa od 3200MT/s. Poslužitelj je opremljen sa 8 DDR4 memorijskih modula kapaciteta 16GB radne memorije (**128 GB ukupno**). Poslužitelj je opremljen sa jednim lokalnim **NVMe SSD-om** kapaciteta **1,92 TB** i sa dodatnih **2 NVMe SSD-a** kapaciteta **7,68 TB** svaki.

CPU radni poslužitelj

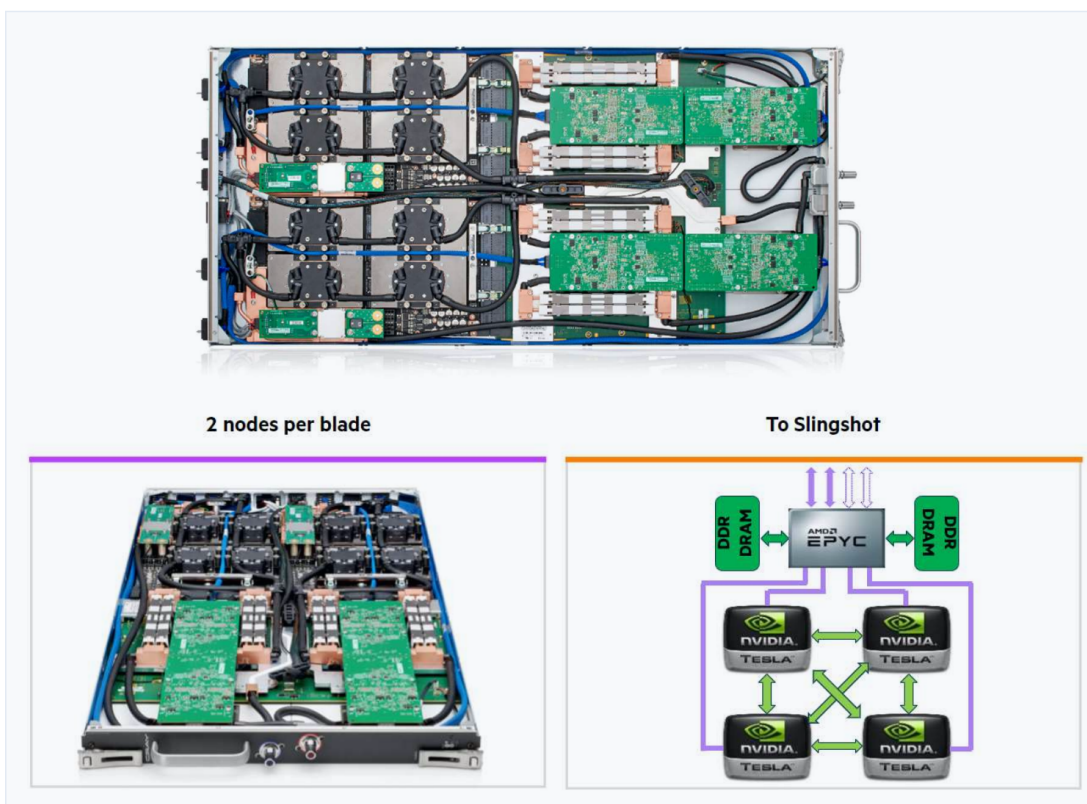
HPE Cray EX425 bladeovi služe za smještaj procesorskih računalnih čvorova ili poslužitelje. Svaki *blade* sadrži 4 računalna poslužitelja. Klaster Supek sadrži ukupno 52 CPU poslužitelja smještenih u 13 računalnih *bladeova*.



Svaki CPU poslužitelj sastoji se od **2 AMD EPYC 7763** procesora sa 64 jezgre (**128 jezgri po poslužitelju** odnosno **6656** na svim CPU poslužiteljima), matične ploče sa 8 memorijskih utora po procesoru maksimalne brzine prijenosa od 3200MT/s i ukupno 16 DDR4 memorijska modula od 16GB radne memorije (**256GB ukupno**).

GPU radni poslužitelj

HPE Cray EX235n bladeovi sadrže po dva GPU poslužitelja. Na klasteru se nalazi ukupno 20 GPU poslužitelja u 10 računalnih *bladeova*.



Blade GPE Cray EX235

Svaki čvor sadrži **1 AMD EPYC 7763** procesor sa 64 jezgre i **4 grafička procesora nVidia HGX A100 u izvedbi SXM** sa **40 GB** ugrađene memorije. GPU poslužitelji opremljeni su sa matičnom pločom sa 8 memorijskih utora po procesoru maksimalne brzine prijenosa od 3200MT/s i sa 8 DDR4 memorijska modula kapaciteta 64GB radne memorije (**512GB ukupno**).

Poslužitelji s velikim memorijskim kapacitetom

Supek sadrži **2 poslužitelja HPE Proliant DL385 Gen 10 Plus v2**. Svaki od poslužitelje opremljen je sa **2 AMD EPYC 7763** procesora. Matična ploča poslužitelja sadrži 16 memorijskih utora po procesoru i podržava brzinu prijenosa od 3200MT/s. 32 DDR4 memorijskih modula kapaciteta **128GB** radne memorije, osiguravaju svakom čvoru ukupno **4096 GB** radne memorije. Poslužitelj je opremljen sa lokalnim **NVMe SSD-om** kapaciteta **1,92 TB**.

Međuveza

Procesorski i spremišni dio superračunala Supek povezani su najnovijom inačicom Crayeve međuveze **Slingshot**, namijenjenoj tipičnim radnim opterećenjima u HPC okruženjima fokusiranim na obradu i razmjenu velike količine podataka.

Glavna stavka u performansama **Slingshota** je brzina od 200 Gbps koja u kombinaciji s 64 porta dostupnih na svakom **preklopniku** i mrežnom topologijom **Dragonfly** postiže iznimne brzine prijenosa.

Osim visokih performansi na nivou hardvera, **Slingshot** međuveza optimizira mrežni prijenos podataka prilagodljivim usmjeravanjem paketa (s obzirom na opterećenje ostalih preklopnika) i kontrolom zagušenja (pojačanjem ili smanjivanjem propusnosti poslužitelja koji generiraju promet).

blocked URL

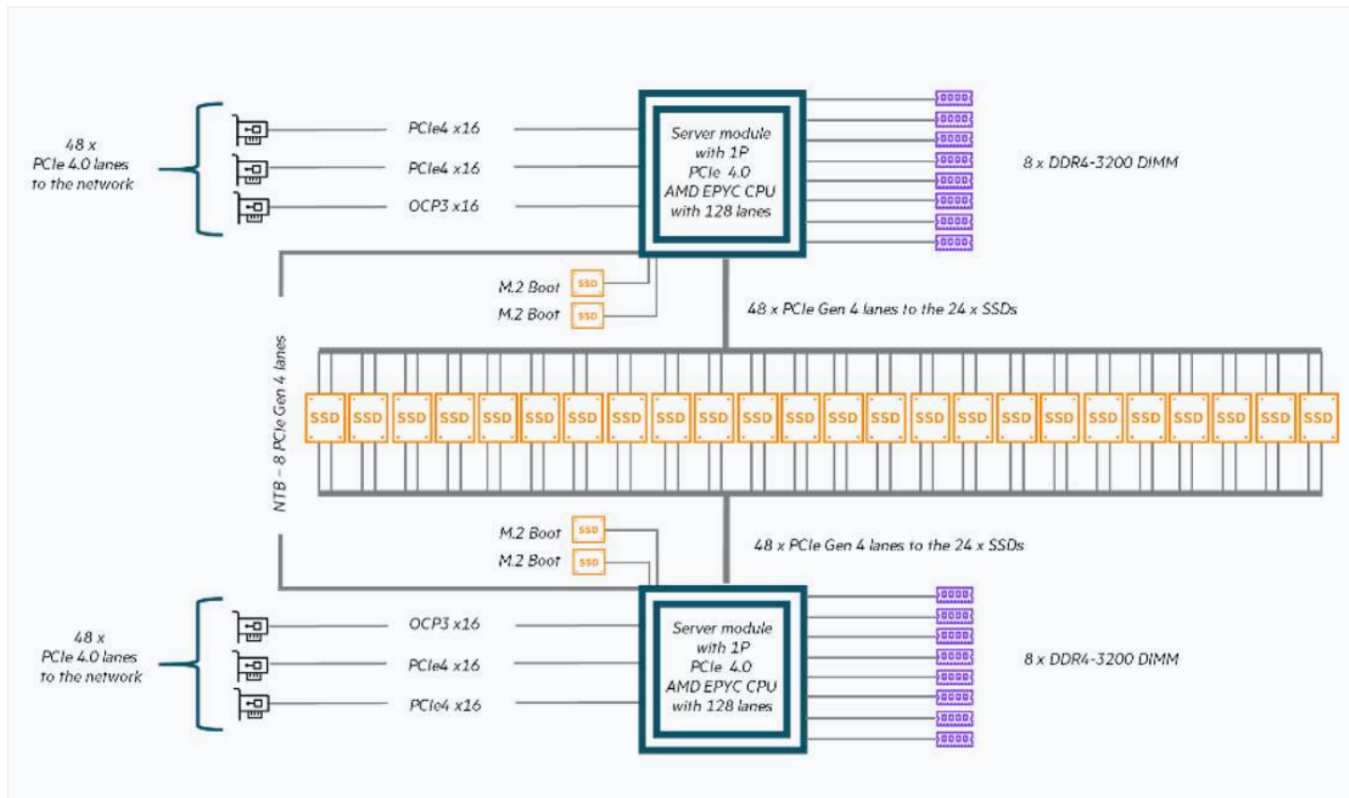
Mrežna topologija Dragonfly: Čvorovi (kružići) su putem preklopnika (kvadratići) povezani u lokalne grupe (putem električnih signala), a grupe čvorova putem globalnih veza (optički kablovi) u mrežu (izvor)

Podatkovno spremište

Podatkovno spremište namijenjeno za potrebe korisničkih poslova za klastera Supek je **HPE ClusterStor E1000**.

Arhitektura za pohranu HPE ClusterStor E1000 temelji se na fleksibilnom dizajnu hardvera koji koristi najnoviju tehnologiju za pohranu uključujući PCI Gen 4 NVMe flash SSD, HPE-ove mrežnu tehnologiju Slingshot 200 Gbps kućišta vrlo visoke gustoće. To omogućuje pohranu s optimiziranim razinom meta podataka zasnovanom na NVMe flashu visokih performansi kako bi se osigurao optimalni I/O put za radna opterećenja koja imaju sekvencijalni I/O ili visoke IOPS-ove, ili oboje. Rješenje omogućuje neovisno skaliranje performansi i kapaciteta.

Osnovi gradivni element temeljen je na računalnom poslužiteljskom čvoru koji osim što služi kao glavni čvor za pohranu kapaciteta (Flash Scalable Storage Units, SSU-F), koristi se i kao poslužiteljski čvor za meta podatke (Metadata Management Unit, MDU) i poslužiteljski čvor za upravljanje (System Management Unit, SMU). Svaki element sadrži 2 matične ploče sa jednim procesorom i 24 mjesta za NVMe SSD pogone koji se popunjavaju ovisno o ulozi elementa.



Shematski prikaz elementa sustava HPE ClusterStor E1000

Rješenje se sastoji od:

1x System Management Unit, SMU, 5x 1,6TB NVMe SSD

1x Metadata Management Unit, MDU, 24x 3,2TB NVMe SSD

6x Flash Scalable Storage Units, SSU-F, 24x 7,68TB NVMe SSD

2x Local Management Network switch, LMN

2x Slingshot switch, SS

1x Flash Scalable Storage Unit, SSU-F, 24x 7,68TB NVMe SSD ukupnog kapaciteta 184TB za lokalni diskovni prostor CPU i GPU poslužitelja