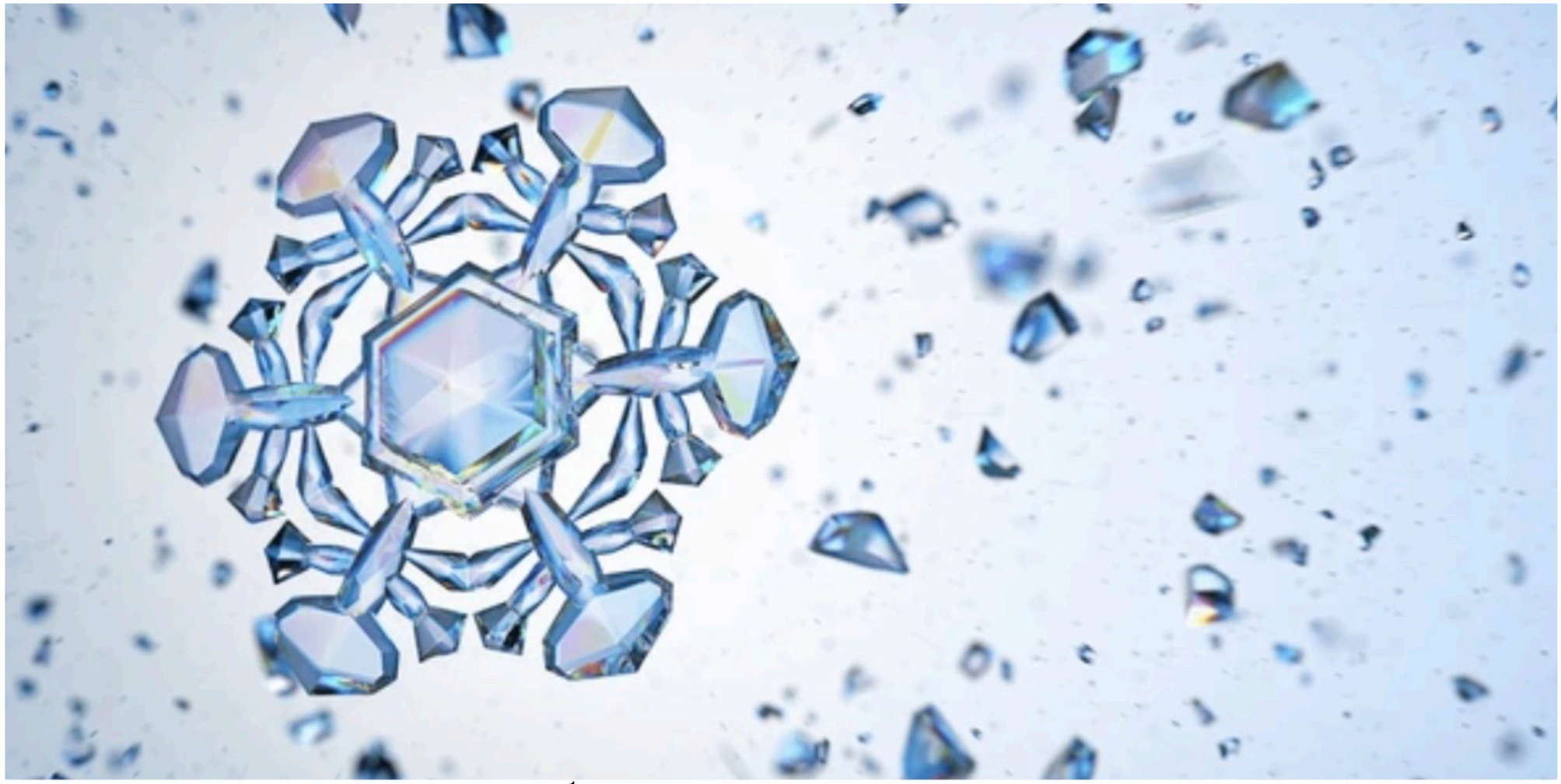
LUMI

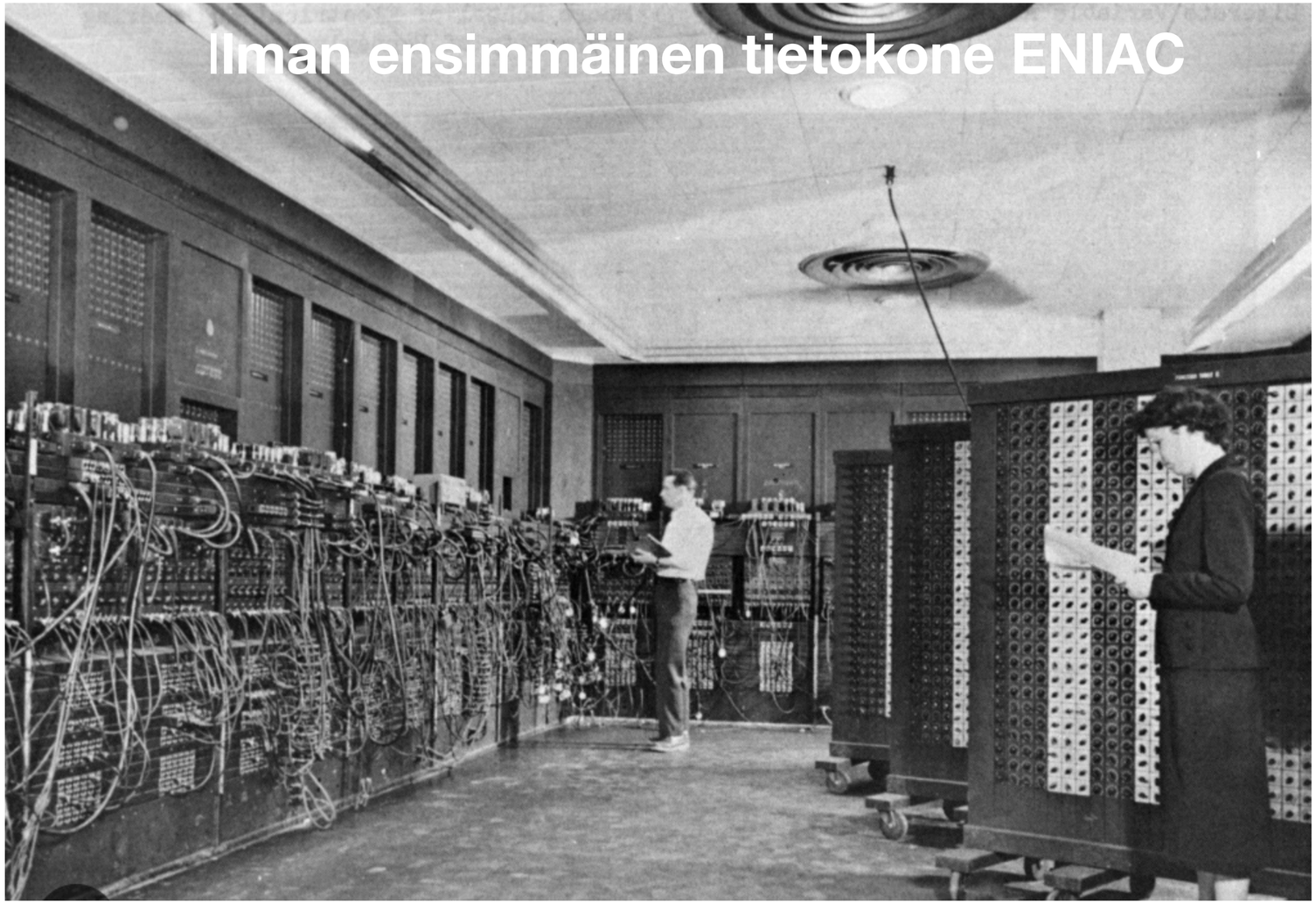
Ja muut Suomessa olevat supertieto-koneiden tarinat

Kristian Salmi
11.12.2024




LUMI

Supertietokoneiden historia



Mitkä ovat maailman tehokkaimmat tietokoneet

Järjestyksessä:

- Vuonna 2016 TOP500-listan ykkössijalla oli kiinalainen Sunway TaihuLight (93 PFLOPS).

- Marraskuussa 2018 ykkössijalla oli Summit (200 PFLOPS).

- Kesäkuussa 2020 ykköspaikan otti Fugaku (415,5 PFLOPS).

- Avattiin virallisesti 2021.LUMI:n teho on yli (550 PFLOPS),

- Toukokuussa 2022 nopeimmaksi nousi ensimmäinen eksaluokan supertietokone Aurora.

- - Aurora on yksi maailman ensimmäisistä eksakokoisista supertietokoneista, joka pystyy suorittamaan yli kvintiljoonaa laskutoimitusta (10 potenssiin 30 FLOPS) sekunnissa.

* **FLOPS** on tietotekniikassa käytetty lyhenne sanoista **FL**oating point **O**perations **P**er **S**econd. Flops mittaa tietokoneen suorituskykyä erityisesti tieteellisessä laskuissa

LUMI-tietokoneen yleiskatsaus

Johdanto

LUMI = (Large Unified Modern Infrastructure) on yksi maailman tehokkaimmista supertietokoneista, joka sijaitsee **Suomessa**. Se on suunniteltu erityisesti tutkimusta ja innovaatioita varten, ja se tarjoaa valtavat laskentatehot eri tieteellisten ja teollisten sovellusten käyttöön. LUMI:n kehittämiseen on osallistunut useita organisaatioita, ja se on tärkeä osa Euroopan Unionin tutkimusinfrastruktuuria.

LUMI:n tausta

- **Perustaminen**: LUMI:n kehittämisprojekti alkoi 2019, ja se avattiin virallisesti **2021**. Se on osa Euroopan Laskentaverkkoa (EuroHPC), joka pyrkii yhdistämään eri maiden laskentaresurssit.
- **Sijainti**: LUMI sijaitsee **CSC - Tieteen tietotekniikan keskuksessa** kansallisessa tutkimuslaitoksessa, joka on yksi *Euroopan johtavista tietotekniikan ja laskennan tutkimuskeskuksista.*

LUMI:n pintapuoliset tekniset tiedot

Laitteisto

- **Prosessorit**: LUMI käyttää **AMD EPYC** -prosessoriteknologiaa, joka mahdollistaa suuren laskentatehon ja energiatehokkuuden.
- **Grafiikkaprosessorit**: Siinä on myös valtava määrä **NVIDIA A100** -grafiikkaprosessoreita, jotka tukevat syvää oppimista ja muita laskentatehoja vaativia sovelluksia.
- **Muisti**: Supertietokoneessa on **yli 500 teratavua muistia**, mikä mahdollistaa suurten datamäärien käsittelyn tehokkaasti.

Suorituskyky

- **Flops**: LUMI:n teho on yli **550 petaflopsia**, mikä tekee siitä yhden maailman nopeimmista supertietokoneista.
- **Energiatehokkuus**: LUMI *on suunniteltu olemaan energiatehokas. Sen käyttö on optimoitu siten, että se vähentää hiilidioksidipäästöjä ja energiankulutusta.*

LUMI:n Käyttötarkoitukset

LUMI:a käytetään laajalti eri tieteellisten alojen tutkimuksessa, mukaan lukien:

Tieteellinen tutkimus

- **Ilmastotiede**: Ilmastonmuutoksen mallintaminen ja ennustaminen.
- **Biotieteet**: Genomiikan ja proteomiikan tutkimus.
- **Fysiikka**: Perusfysiikan ja astrofysiikan simulaatiot.
- **Kemia**: Kemiallisten reaktioiden simulointi ja materiaalitutkimus.

Teollisuus

- **Lääketiede**: Uuden lääkkeen kehittäminen ja kliinisten kokeiden simulointi.
- **Energia**: Uusien energiateknologioiden tutkiminen ja kehittäminen.
- **Automaatio**: Teollisen automaation ja robotiikan kehittäminen.

LUMI:n rooli Euroopassa

LUMI on keskeinen osa Euroopan supertietokoneteollisuutta ja infrastruktuuria. Se mahdollistaa:

- **Yhteistyö**: Euroopan eri maiden tutkimuslaitokset voivat käyttää LUMI:a, mikä edistää kansainvälistä yhteistyötä.
- **Innovaatio**: Supertietokoneen tarjoamat resurssit tukevat innovaatioita ja uusien teknologioiden kehittämistä.
- **Koulutus**: LUMI tarjoaa myös koulutusmahdollisuuksia tutkijoille ja opiskelijoille, mikä edistää tieteen ja teknologian kehitystä.

Tulevaisuuden näkymät

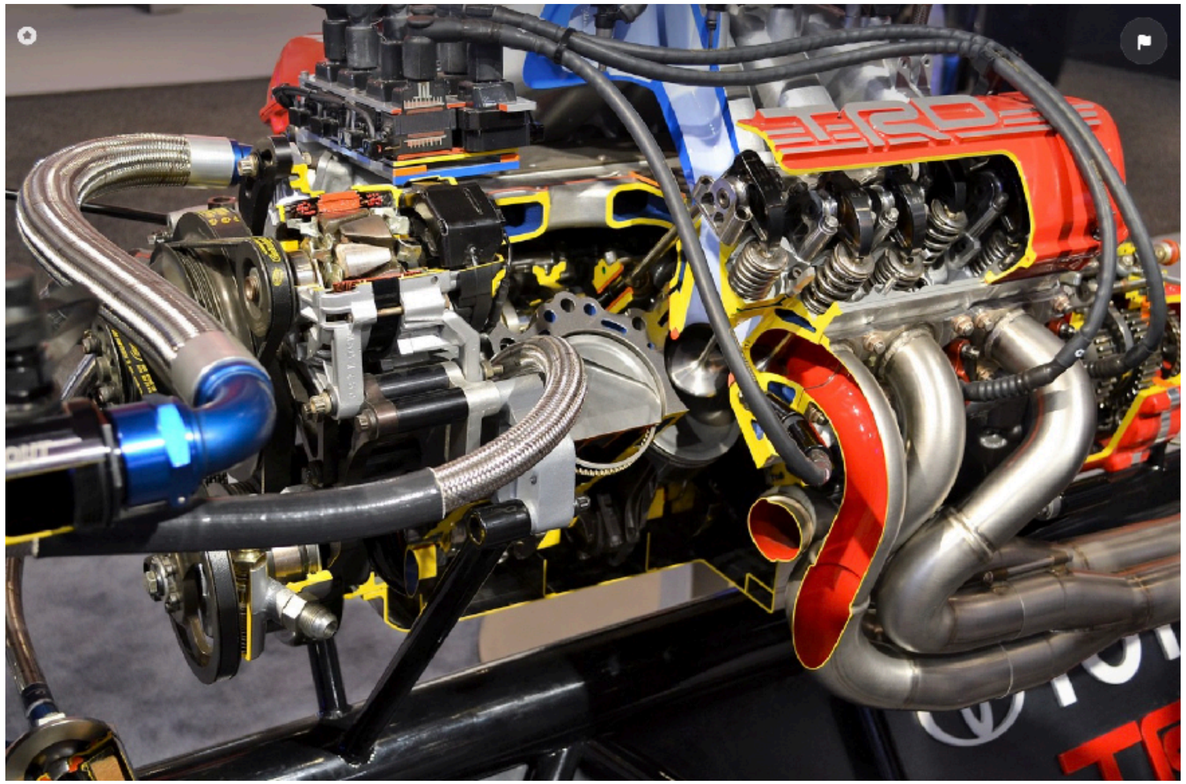
LUMI:n tulevaisuus näyttää lupaavalta. Sen käyttö laajenee jatkuvasti, ja uusia tutkimusprojekteja käynnistyy.

Tavoitteena on:

- **Kestävyys**: Kehittää entistä energiatehokkaampia ja kestävämpiä laskentaratkaisuja.
- **Innovaatio**: Edistää innovaatioita eri aloilla, erityisesti terveydenhuollossa ja ympäristötutkimuksessa.
- **Yhteistyö**: Vahvistaa kansainvälistä yhteistyötä ja verkostoitumista eri tutkimuslaitosten välillä.

LUMI tietokoneesta lähemmin
…
(Mitä salaisuuksia sieltä löytyy?)

LUMI:n osat
(Kurkistus konepellin alle)



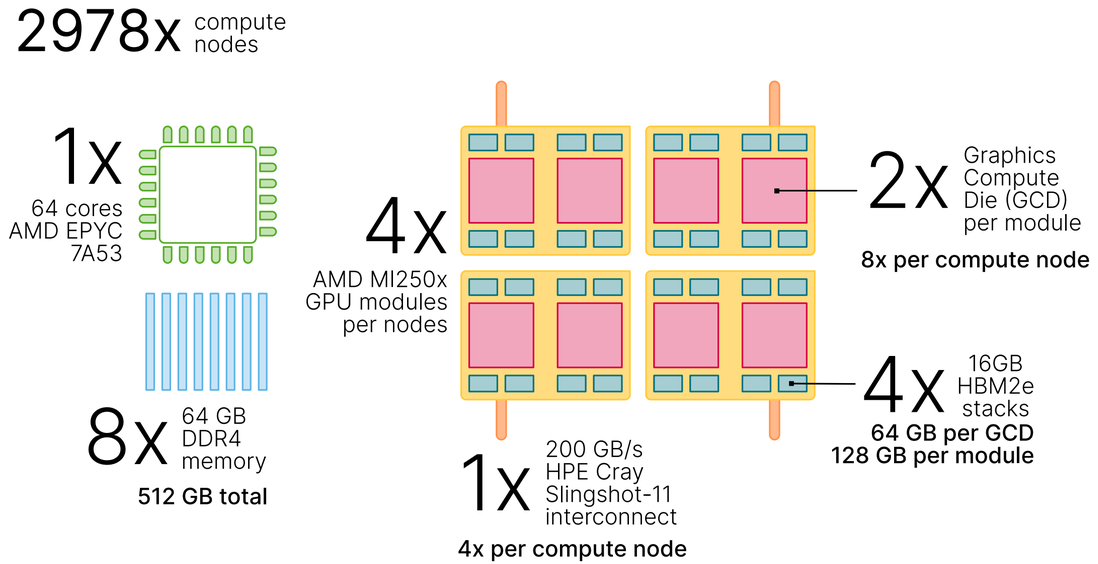
LUMI tietokoneen laskentatehot….

- LUMI:n ensisijainen laskentateho löytyy LUMI-G-laitteisto-osiosta, jossa on GPU-kiihdytetyt solmut, jotka käyttävät **AMD Instinct MI250X** grafiikkasuorittimia.
- Tämän lisäksi tarjolla on pienempi LUMI-C-suorittimia sisältävä laitteisto-osio, jossa on **AMD EPYC "Milan"-suorittimet**,
- sekä pieni LUMI-D-data-analytiikkalaitteisto-osio, jossa on suuria muistisolmuja **(4 TB) ja NVIDIA A40.** -> GPU:t tietojen visualisointiin.

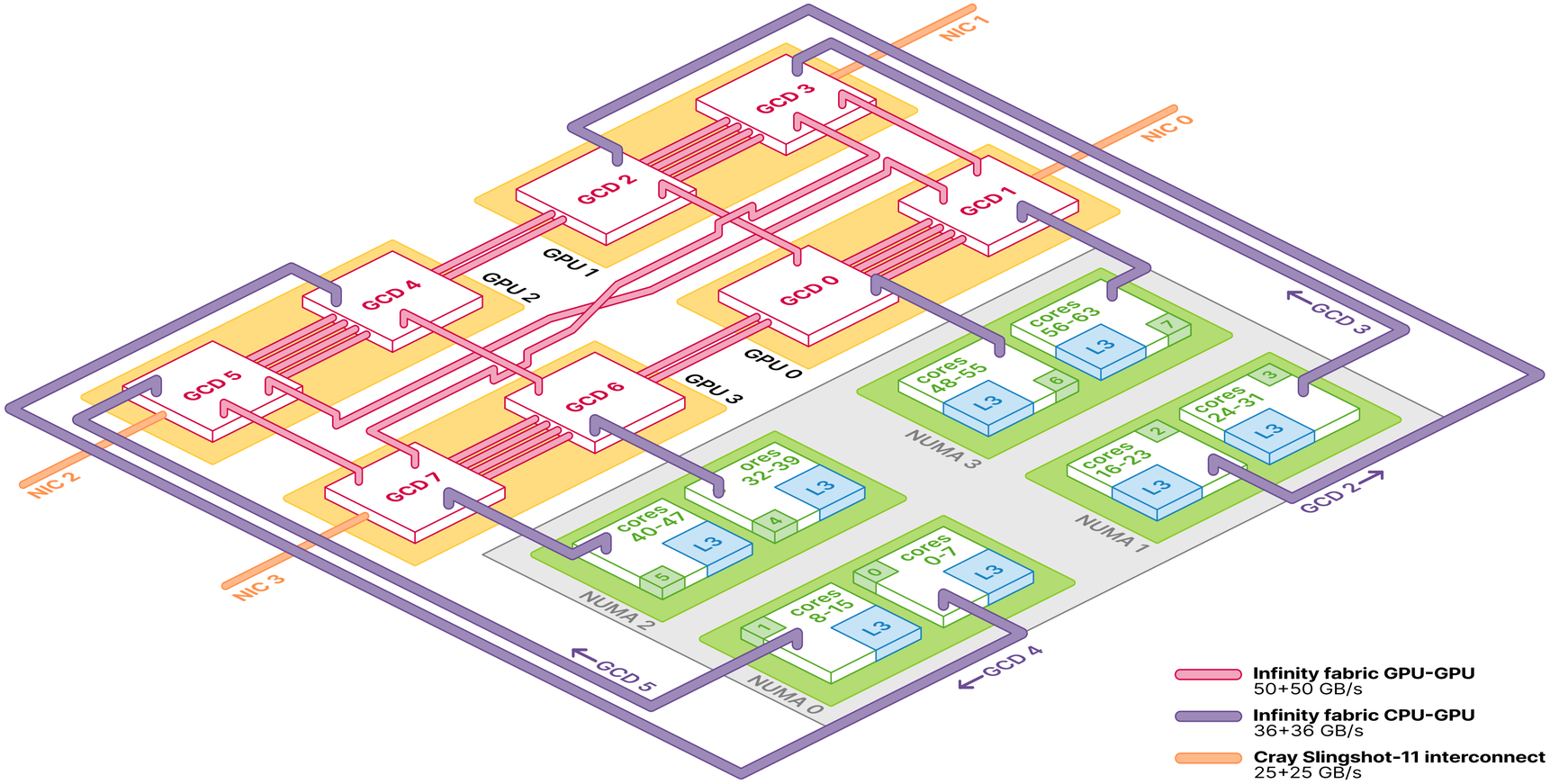
LUMI G

GPU-solmut - LUMI-G¶

- GPU = Graphical Processing unit
- LUMI-G-laitteistoosio koostuu 2978 solmusta,
- joissa on 4 AMD MI250x grafiikkasuoritinta
- ja yksi 64-ytiminen AMD EPYC "Trento" -suoritin. LUMI-G:n yhteenlaskettu HPL Linpack -suorituskyky on 379,70 PFlop/s.



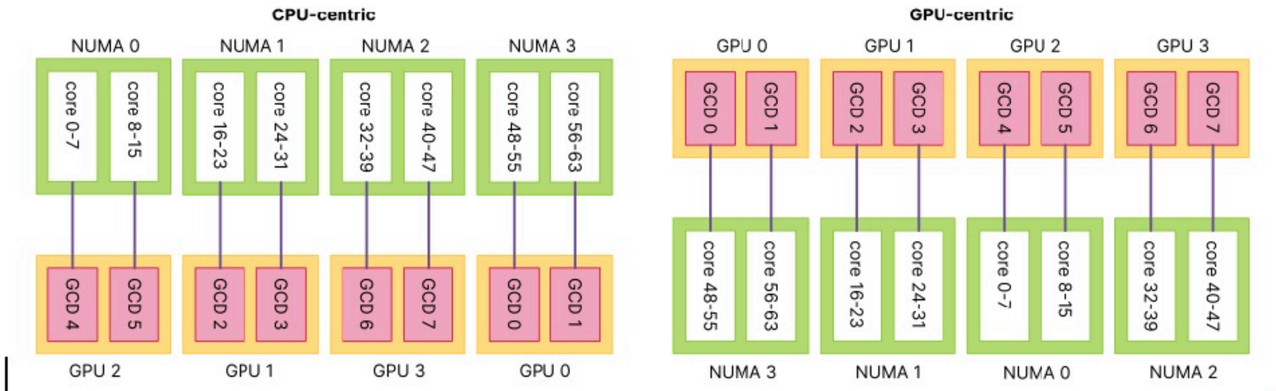
Yleinen toimintajohteinen kuva LUMI G lohkosta



GPU ja NUMA solmut

Non-uniform memory access, or NUMA,

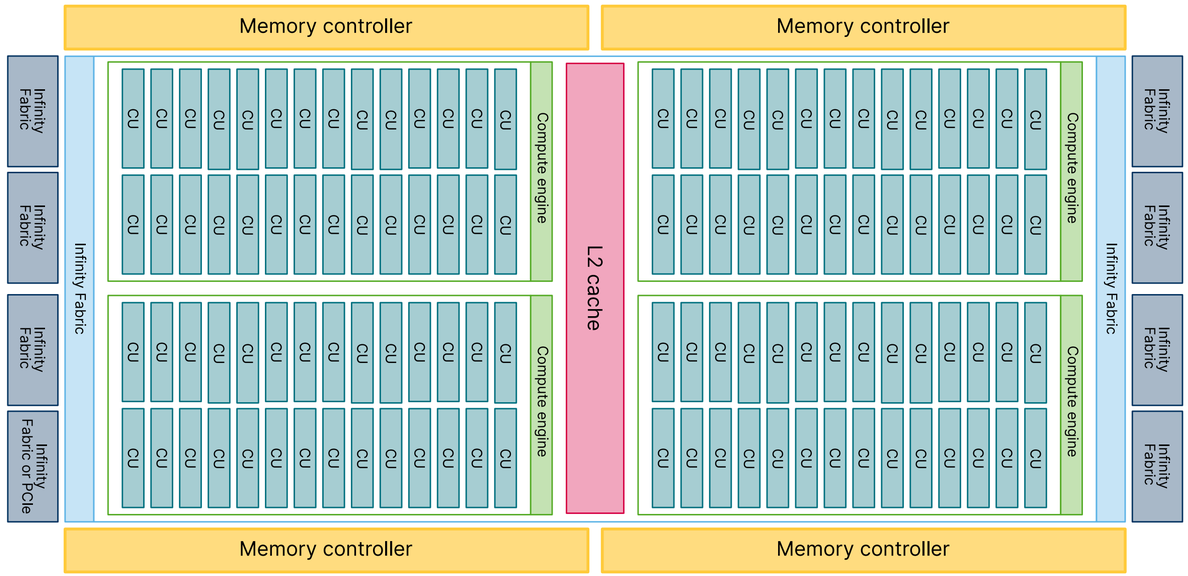
- On tärkeää huomata, että GPU (GCD) -numeroinnin ja **NUMA**-solmunumeron välillä ei ole suoraa korrelaatiota. Esimerkiksi NUMA 0/ core 0 ei ole yhdistetty GPU 0:aan/GCD 0:aan. NUMA-solmun oikea sitominen GPU:hun saattaa olla ratkaisevan tärkeää sovelluksesi optimaalisen suorituskyvyn saavuttamiseksi. Vieressä oleva kuva havainnollistaa näitä linkkejä sekä CPU- että GPU-näkökulmasta.
- **Epäyhtenäinen muistin käyttö eli NUMA** on menetelmä mikroprosessorien klusterin (ryppään) määrittämiseksi moniprosessointijärjestelmässä, jotta ne voivat jakaa muistia paikallisesti. Ajatuksena on parantaa järjestelmän suorituskykyä ja antaa sen laajentua käsittelytarpeiden kehittyessä.



CPU-GPU links from a CPU centric or GPU centric point of view

L2 välimuisti

- L2-välimuisti parantaa myös synkronointiominaisuuksia **algoritmeille**, jotka luottavat atomioperaatioihin koordinoidakseen viestintää koko GPU:n välillä. Nämä atomioperaatiot suoritetaan lähellä L2-välimuistin muistia.



CPU nodes - LUMI-C

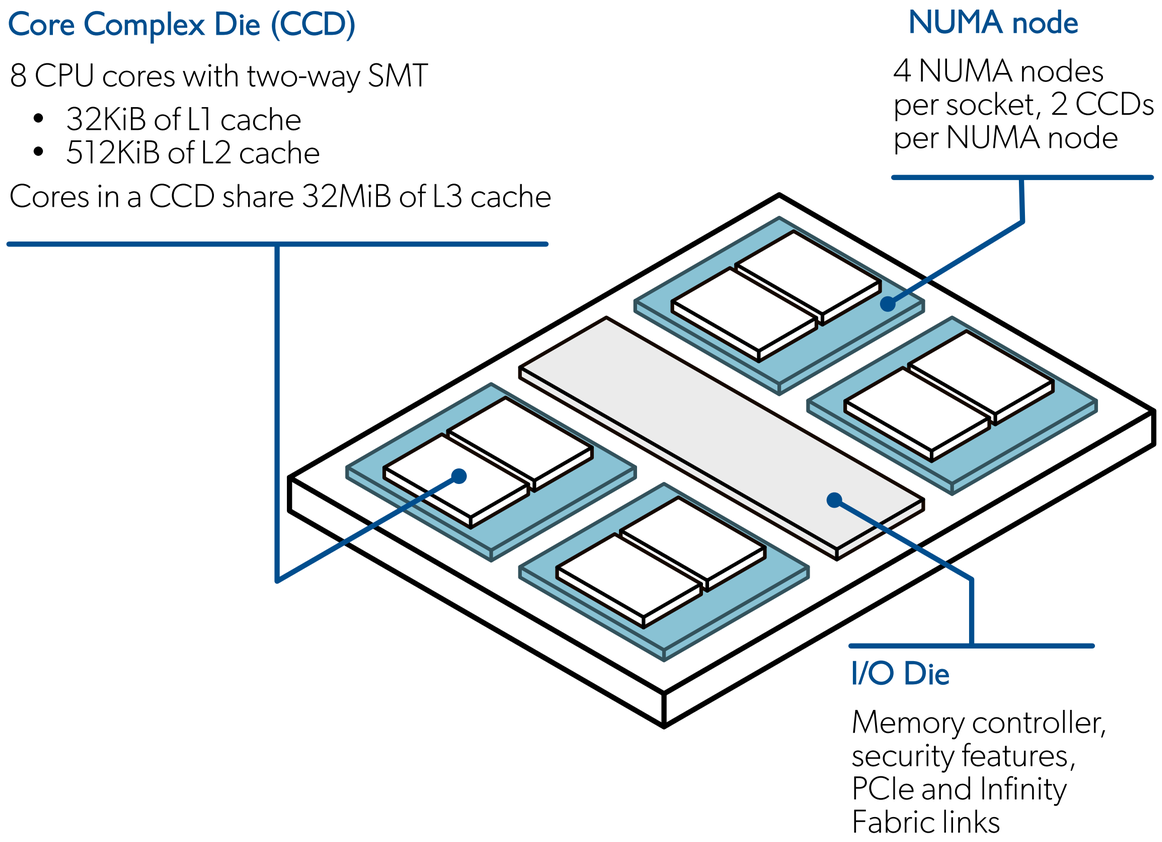
- LUMI-C-laitteisto-osio koostuu 2048 CPU-pohjaisesta laskentasolmusta. Jotkut näistä solmuista sisältävät enemmän muistia kuin toiset vieressä olevan taulukon mukaisesti.

Nodes	CPUs	CPU cores	Memory	Disk	Network
1888	2x AMD EPYC 7763 (2.45 GHz base, 3.5 GHz boost)	128 (2x64)	256 GiB	none	1x 200 Gb/s
128	2x AMD EPYC 7763 (2.45 GHz base, 3.5 GHz boost)	128 (2x64)	512 GiB	none	1x 200 Gb/s
32	2x AMD EPYC 7763 (2.45 GHz base, 3.5 GHz boost)	128 (2x64)	1024 GiB	none	1x 200 Gb/s

LUMI CPU

- Kukin LUMI-C-laskentasolmu on varustettu kahdella AMD EPYC 7763 -suorittimella, joissa kussakin on 64 ydintä ja jotka toimivat 2,45 GHz:n taajuudella, eli yhteensä 128 ydintä solmua kohti. **Ytimet tukevat kaksisuuntaista samanaikaista monisäikeistystä (SMT)**, mikä mahdollistaa jopa 256 säiettä solmua kohti.

- AMD EPYC 7763 -suorittimen ytimet ovat **"Zen 3" -laskentaytimiä, jotka ovat samat kuin Ryzen 5000 -sarjan kuluttajasuorittimissa.** Nämä ytimet ovat täysin x86-64-yhteensopivia ja tukevat AVX2:n 256-bittisiä vektoriohjeita 16 kaksinkertaisen tarkkuuden FLOP/kellon maksimiläpäisykyvyn saavuttamiseksi (AVX2 FMA -toiminnot). Ytimessä on 32 KiB yksityistä L1-välimuistia, 32 KiB käskyvälimuistia ja 512 KiB L2-välimuistia.



Mihin tietoja tallennetaan?

- Jokaisella käyttäjällä on kotihakemisto ($HOME), Se on projektikohtainen, joka voi sisältää jopa 20 Gt tietoa.
- Se on tarkoitettu käyttäjien asetustiedostojen ja henkilötietojen tallentamiseen.
- Käyttäjän kotihakemisto tyhjennetään, kun projekti loppuu tai sitä ei käytetä aktiivisesti
- LUMI-verkkotiedostojärjestelmän levytallennusalueet viereisessä taulukossa
- LUMIssa on useita verkkopohjaisia levytallennusalueita.
- Vieressä olevassa taulukossa on yleiskatsaus.

	Path	Intended use	Hardware partition used		
User home	/users/<username>	User home directory for personal and configuration files	LUMI-P		
Project persistent	/project/<project>	Project home directory for shared project files	LUMI-P		
Project scratch	/scratch/<project>	Temporary storage for input, output or checkpoint data	LUMI-P		
Project flash	/flash/<project>	High performance temporary storage for input and output data	LUMI-F		
	Quota	Max files	Expandable	Retention	Billing rate
User home	20 GB	100k	No	User lifetime	NA
Project persistent	50 GB	100k	Yes, up to 500GB	Project lifetime	1x
Project scratch	50 TB	2000k	Yes, up to 500TB	90 days	1x
Project fast	2 TB	1000k	Yes, up to 100TB	30 days	10x
	Quota	Max buckets	Max objects per bucket	Retention	Billing rate
Object storage	150 TB	1000	500 000	project lifetime	0.5x

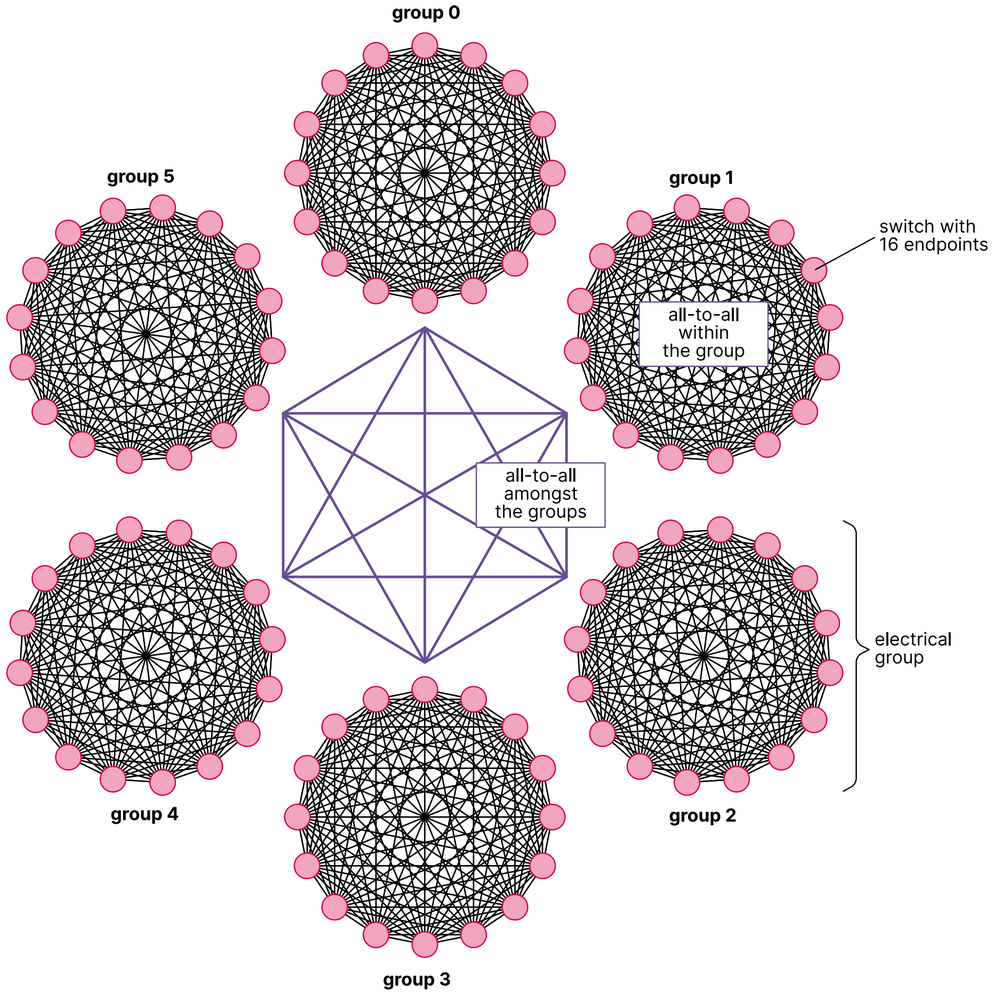
Verkko ja yhteenliittäminen

- Kaikki LUMI-laskentasolmut käyttävät HPE **Cray Slingshot-11 200 Gbps** verkkoyhteyttä (NIC).
- LUMI-C-solmuissa (CPU) on **yksi päätepiste,** kun taas
- LUMI-G-solmuissa (GPU-solmuissa) **on neljä päätepistettä** – yksi jokaiselle AMD MI250x GPU -moduulille.
- Jokainen päätepiste tarjoaa jopa 50 Gt/s kaksisuuntaista kaistanleveyttä. HPE Cray Slingshot NIC sisältää korkean suorituskyvyn RDMA- ja laitteistokiihdytyksen MPI- ja SHMEM-pohjaisille ohjelmistoille.

Cray Slingshot-11 200 Gbps

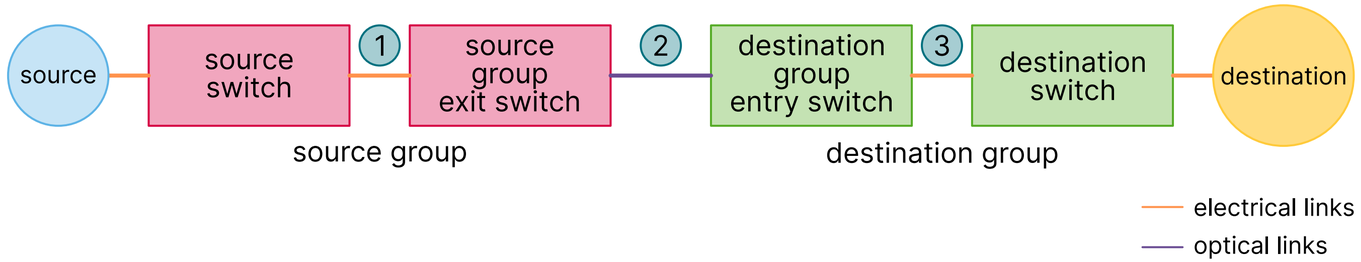
Verkko-topologia

- LUMI käyttää (dragonfly) sudenkorentoverkkotopologiaa (**tämä ei ole MESH?**).
- Tässä topologiassa joukko päätepisteitä, esimerkiksi laskentasolmut, on kytketty kytkimeen.
- Useita kytkimiä, jotka on yhdistetty kaikkiin ryhmän luomiseksi.
- Tätä ryhmää kutsutaan joskus sähköryhmäksi, koska nämä kytkimet voidaan yhdistää lyhyillä kuparikaapeleilla.
- Sähköiset ryhmät yhdistetään sitten toisiinsa täydellisesti.
- Optisia kaapeleita käytetään ryhmien väliseen viestintään, koska etäisyydet ovat paljon suuremmat.
- Vieressä olevassa kuvassa on graafinen esitys sudenkorentotopologiasta.



Yhteenvetona LUMI-verkoston sudenkorento voidaan tiivistää seuraavasti:

- **Taso 1:** useita laskentasolmuja, jotka on yhdistetty kytkimeen
- **Taso 2:** useita kytkimiä, jotka on yhdistetty kuparikaapeleilla muodostaen sähköryhmän
- **Taso 3:** useita sähköryhmiä, jotka on yhdistetty optisilla kaapeleilla
- Sudenkorentotopologia mahdollistaa viestinnän suorittamisen enintään 3 Taso-hyppelyssä:
- yksi hyppy l**ähderyhmän sisällä,**
- yksi hyppy **ryhmän välillä** ja
- yksi hyppy **kohderyhmän sisällä**.



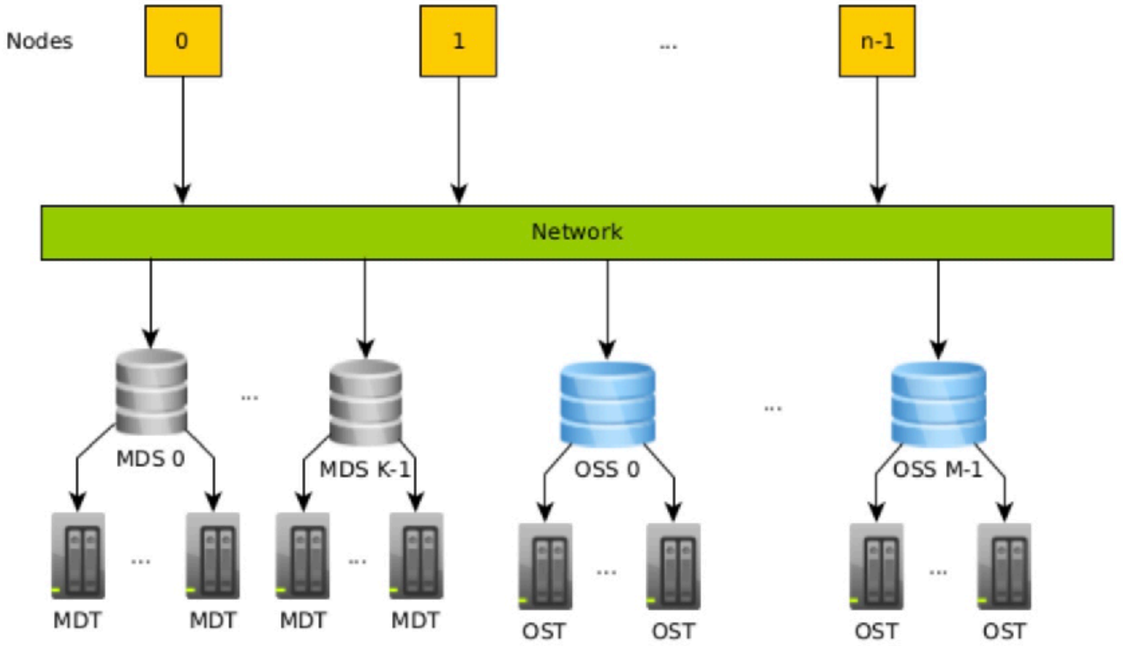
LUMI-TIETOKONEEN OSIOT

- **LUMI-G:** AMD MI250x GPU -solmut
- **LUMI-C:** CPU-solmut
- **LUMI-M:** kirjautumissolmut, LUMI-D, suuret muistisolmut ja hallintatelineet
- **LUMI-P:** neljä mekaanista levyä käyttävää Lustre-tiedostojärjestelmää
- **LUMI-F:** Lustre-tiedostojärjestelmä, joka käyttää Flash-pohjaista tallennustilaa
- **LUMI-C:** CPU-osio LUMI-C sisältää 8 sähköistä ryhmää, joissa on 256 solmua. Ryhmät koostuvat 16 kytkimestä, jotka on kytketty kaikkiin. Jokaiseen kytkimeen on kytketty 16 solmua.

	LUMI-G	LUMI-C	LUMI-M	LUMI-F	LUMI-P
LUMI-G	276 TB/s	38.4 TB/s	2.4 TB/s	7.2 TB/s	9.6 TB/s
LUMI-C		22.4 TB/s	0.8 TB/s	3.2 TB/s	3.2 TB/s
LUMI-M				0.1 TB/s	0.4 TB/s
LUMI-F					0.4 TB/s

Lustre-tiedostojärjestelmänäkymä

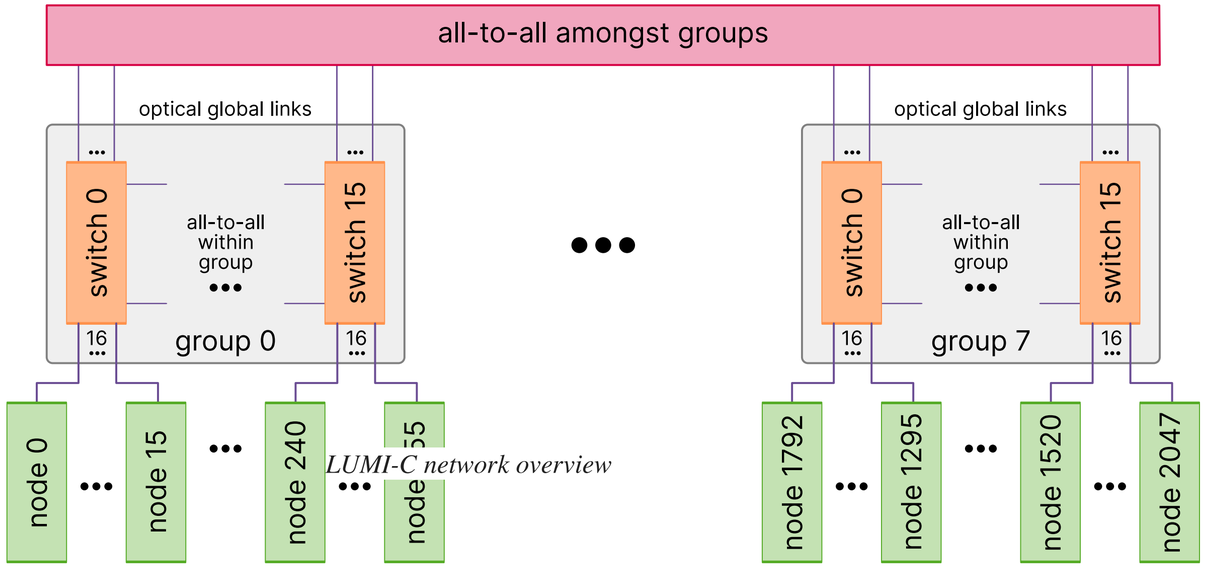
- Lustre on suunniteltu tehokkaaseen rinnakkaiseen I/O-liitäntään suurille tiedostoille.
- Kuitenkin, kun käsitellään pieniä tiedostoja ja intensiivisiä metatietotoimintoja, MDS/MDT voi muodostua pullonkaulaksi. Esimerkiksi kun käyttäjä avaa/sulkee tiedoston monta kertaa silmukassa, MDT:n työmäärä kasvaa.
- Kun useat käyttäjät tekevät samanlaisia toimintoja, metatietotoiminnot voivat hidastaa koko järjestelmää ja vaikuttaa moniin käyttäjiin.
- Koska kirjautumis- ja laskentasolmut jakavat tiedostojärjestelmän, tämä voi näkyä jopa tiedostojen hitaana muokkauksena kirjautumissolmussa.
- Lisäksi, jos rinnakkaisessa sovelluksessa eri prosessit suorittavat paljon toimintoja samoilla pienillä tiedostoilla, metatietotoiminnot voivat hidastua.
- Viattomat Linux-komennot voivat myös lisätä metatietojen työmäärää: esimerkiksi ls -l tulostaa tiedostojen metatiedot, ja komennon antaminen hakemistossa, jossa on paljon tiedostoja, aiheuttaa monia pyyntöjä MDS:lle.



LUMI-C-verkon yleiskatsaus

CPU-osio LUMI-C

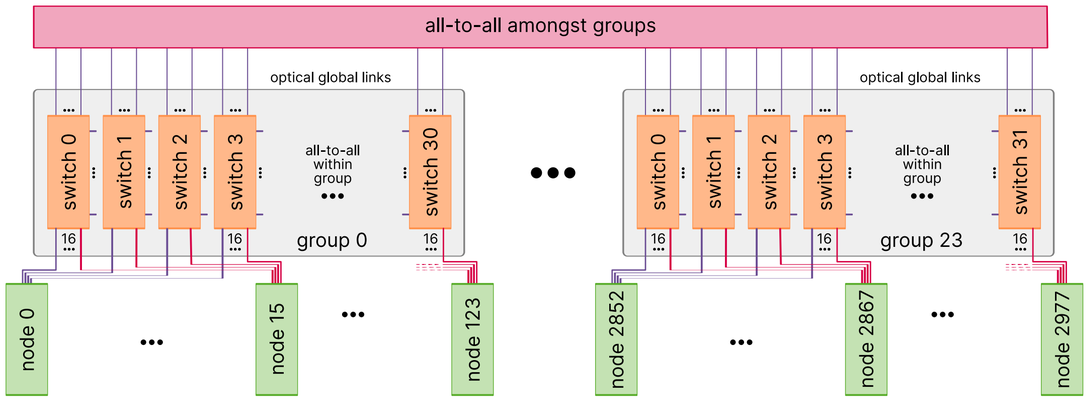
- sisältää **8** sähköistä ryhmää, joissa on **256** solmua.
- Ryhmät koostuvat **16** kytkimestä, jotka on kytketty toisiinsa.
- Jokaiseen **kytkimeen on kytketty 16 solmua**.



LUMI-C network overview

LUMI-G
Yleiskatsaus

- GPU-osiossa LUMI-G on **24 sähköistä ryhmää**, j**oissa on 124 solmua,** paitsi viimeinen ryhmä, joka sisältää 126 solmua.
- Ryhmät koostuvat **32 kytkimestä, jotka on kytketty toisiinsa**.
- **Jokaiseen kytkimeen on kytketty 16 päätepistettä.**
- **LUMI-G-laskentasolmuissa on 4 päätepistettä solmua kohti, joista jokainen on kytketty eri kytkimiin.**



LUMI Yhteenveto

- LUMI-supertietokone on **huipputeknologian merkkipaalu, joka tarjoaa valtavat laskentatehot tieteelle ja teollisuudelle.**
- Sen kehitys ja käyttö ovat merkittäviä askeleita kohti tulevaisuuden innovaatioita ja tutkimusta. LUMI on esimerkki siitä, kuinka huipputeknologia voi tukea kestävää kehitystä ja edistää tieteellistä tutkimusta globaalisti.

Nyt LUMI SAA UUDEN KAVERIN…

ROIHU korvaa MAHDIN ja PUHDIN…

Mikä Roihu?

Marraskuu 20 pv 2024 Helsingin Sanomat:

- Suomi tulee vahvistamaan omaa asemaansa supertietokonemarkkinoilla uudella Roihu-supertietokoneella. Supertietokone tulee perustumaan **Eviden BullSequana XH3000** -järjestelmään.

- Nykyiset Mahdin ja Puhdin korvaava Roihu tulee jopa **kolminkertaistamaan edeltäjiensä laskentatehon yhteensä 486 CPU- ja 132 GPU-laskentasolmun voimin.**

- Suomessa jyllää tällä hetkellä kolme merkittävää supertietokonetta, maailman kahdeksanneksi tehokkain LUMI, Mahti ja Puhti. Siinä missä LUMI on yhteiseurooppalaisen EuroHPC:n, ovat **Mahti ja Puhti, sekä ne korvaava Roihu kansallisia supertietokoneita.** Siinä missä Mahti ylti teoriassa yhteensä 9,5 PetaFLOPSin ja Puhti 1,8 PFLOPSin suorituskykyyn, luvataan uudelle **Roihulle jopa 49 PFLOPSin suorituskykyä.** LUMI-supertietokoneen teoreettinen maksimisuorituskyky on virallisesti 380 PFLOPSia ja teoriassa jopa 550 PFLOPSin luokassa.

ROIHUN TEKNIIKKA JA KOHDENNUKSET

- Roihu perustuu Bull Sequana XH3000 -järjestelmiin ja se tulee sisältämään yhteensä **486 CPU-solmua AMD:n Turin Epyc-prosessoreilla** (Zen 5c) sekä **132 GPU-solmua NVIDIAn GH200 -GPU-kiihdyttimillä.** (*Hybridi*)
- **Kuhunkin CPU-solmuun kuuluu kaksi 192-ytimistä Turinia eli prosessoriytimiä tulee olemaan käytössä yhteensä 186 624 kappaletta, kun kussakin GPU-nodessa on neljä GPU:ta eli yhteensä 528 GH200-kiihdytintä.** (*Hybridi*)
- Järjestelmään kuuluu myös **GPU-erityissolmuja visualisointitarpeisiin sekä CPU-erityissolmuja laajennetulla muistilla, mutta missään ei mainita onko ne laskettu mukaan edellä mainittuihin laskentasolmuihin.** (*Hybridi*)
- Myös kaikkia Suomen supertietokoneita palvelevaa **Allas-datahallintajärjestelmän tallennuskapasiteettia tullaan kasvattamaan samassa yhteydessä.**

ROIHUN KÄYTTÖKOHTEET

- Roihua tullaan hyödyntämään **paitsi tekoälytehtäviin, myös perinteisempään supertietokonelaskentaan, kuten entsyymien ja proteiinien atomitason mallinnuksiin ja ilmastoskenaarioiden pyörittämiseen.**

- Kaikkein suurinta laskentatehoa vaativat tehtävät lasketaan kuitenkin jatkossakin EuroHPC:n LUMI-supertietokoneella. **Roihu tullaan valjastamaan myös korkeakoulujen opiskelijoiden käyttöön suurteholaskentaan perehdyttämiseksi jo opiskeluaikana tulevaisuuden haasteita varten**.

- Roihu tullaan sijoittamaan CSC:n Kajaanin datakeskukseen ja sen käyttöönoton odotetaan tapahtuvan jo vuoden 2025 loppuun mennessä.

Tätä esitystä tehdessä

- On käytetty tekoälyä tiedon keruussa
- Selaimena Opera, sekä ARIA Google searchin päällä.
- Aria on selain-AI, joka on integroitu Opera selaimeen. Se hyödyntää Operan AI-moottoria ja useita suuria kielimalleja. Aria auttaa käyttäjiä ymmärtämällä heidän kysymyksiään, analysoimalla niitä ja vastaamalla inhimillisellä tavalla.

Kiitoksia mielenkiinnosta

- Onko kysymyksiä?
- Onko vastauksia?
- Voidaanko keskustella?