

Liquid trae eficiencia a la UC3M

Eficiencia operativa

2X

Más
Tokens por Watio

Eficiencia de coste

50%

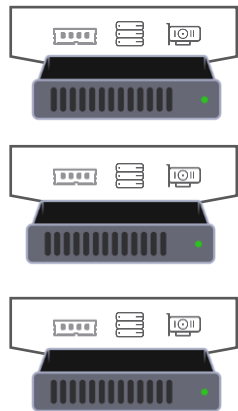
Más
Tokens por Euro

“Energía, no computación, será el embudo #1 para el progreso de la IA”

- Mark Zuckerberg, CEO, Meta

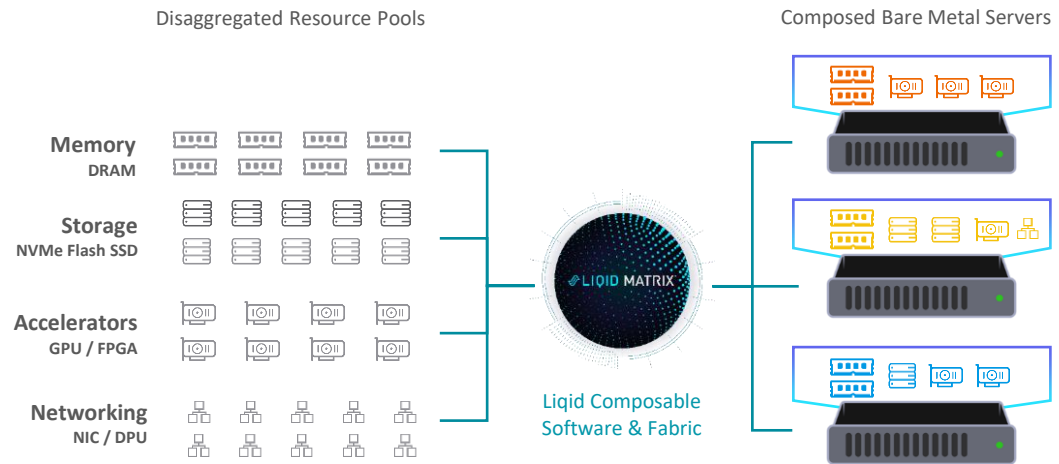
Qué es una infraestructura CDI

Conventional AI Servers



- ! Low Device Density
- ! High Power/TCO
- ! Limited Flexibility

Dynamically Configured AI Servers



- ✓ Highest AI Performance
- ✓ Optimized Efficiency
- ✓ Maximum Flexibility

Los components de Liquid

Devices

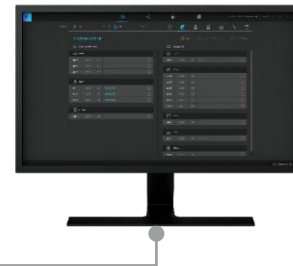
PCIe: GPU, DRAM, FPGA, DPU, etc.
CXL: DRAM



Chasis de expansión



LIQID MATRIX™



Liquid Composable

- Expansion Chassis
- Host Interface Card
- Optical Cables
- Liquid Matrix Software

Host Interface Card

- Gen5 x16 Per Link



Servidores

1U, 2U, 2U4N, Blade, etc.



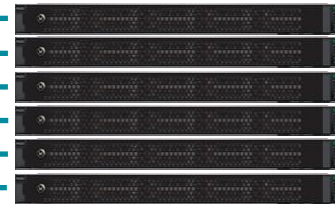
Infraestructura, cómo se integra.

Composable Software



Network Switch

Ethernet | InfiniBand



Servers

1-4U | 4N2U | Blade | Edge



Fabric Switch

PCIe or CXL



Expansion Chassis

PCIe: GPU | FPGA | NVMe | NIC | DPU
CXL: DRAM

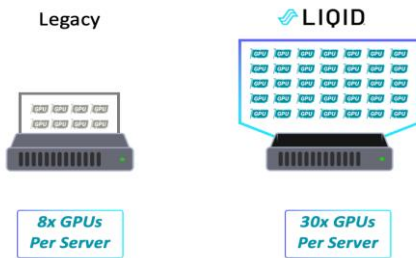
Qué beneficios da Liquid a UC3M

Maximiza Rendimiento

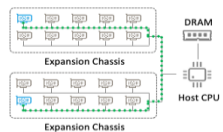


Conecta hasta 30 GPUs/servidor.

Incrementa 30% el rendimiento.

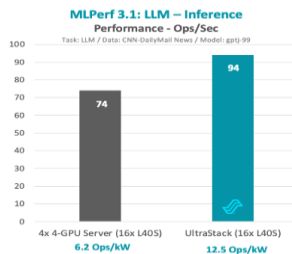


Liquid P2P. Comunicaciones sin pasar por la CPU



3x ancho de banda

1/10 latencia



30% más rendimiento

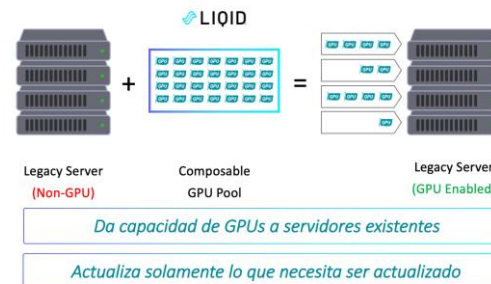
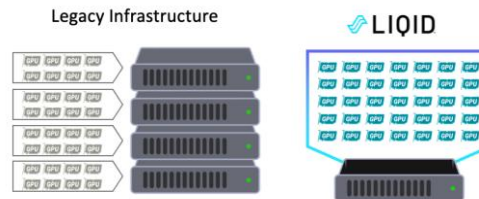
Reduce Costes



Aumenta la eficiencia energética.

Reduce el número de servidores.

Alarga la vida de la infraestructura.



Aumenta la agilidad

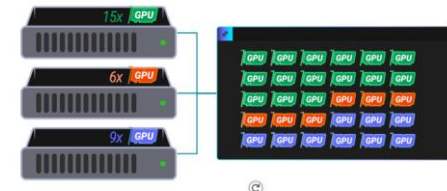


Utiliza cualquier tipo de GPU.

Crece con las necesidades de UC3M.

Tecnología puntera.

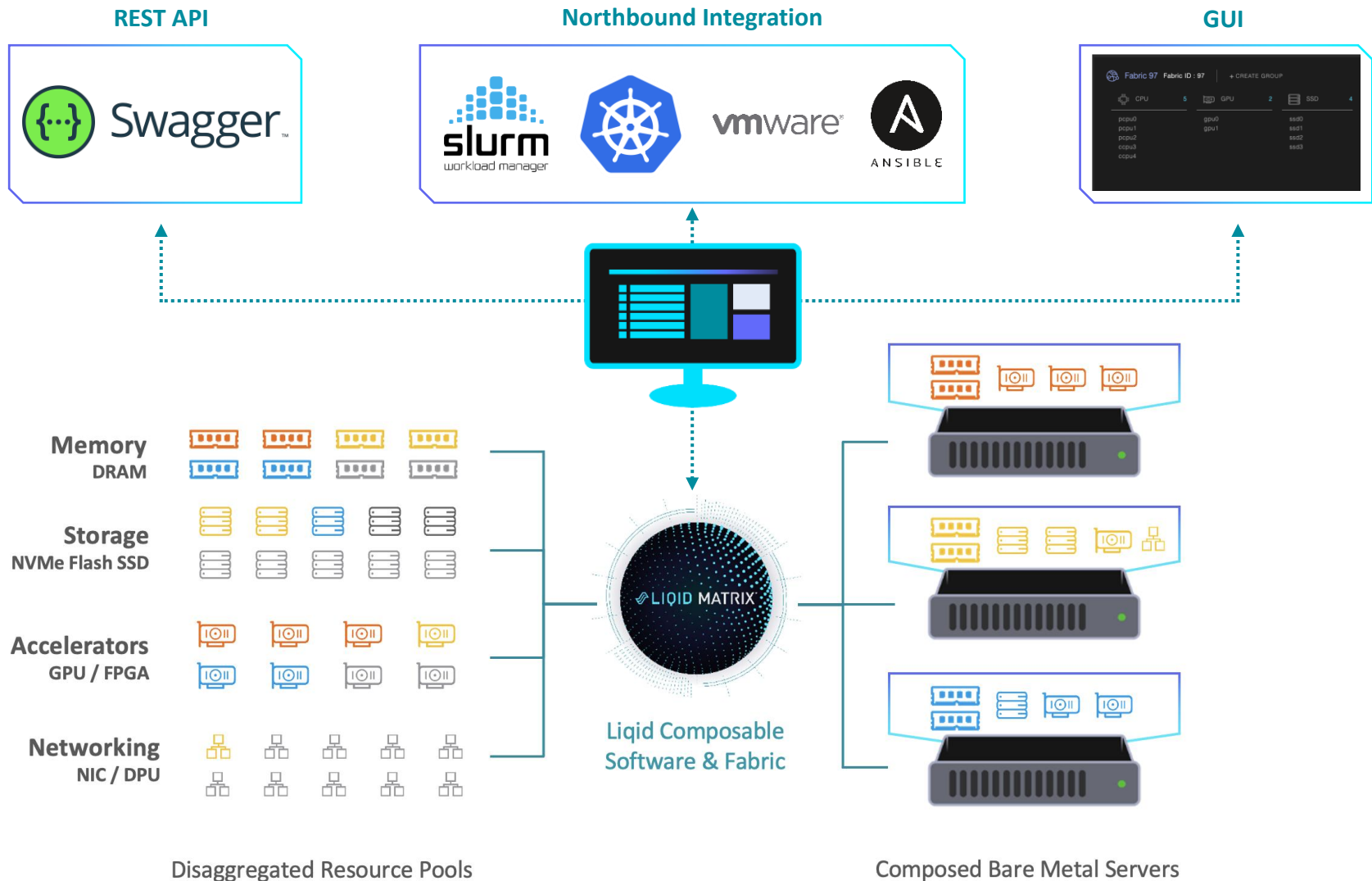
GPUS dinámicas



Qualified GPU Vendors

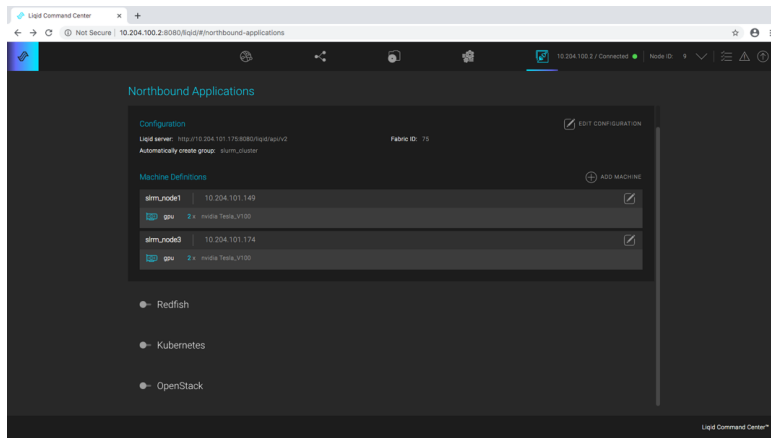


Un software que se integra con lo existente.



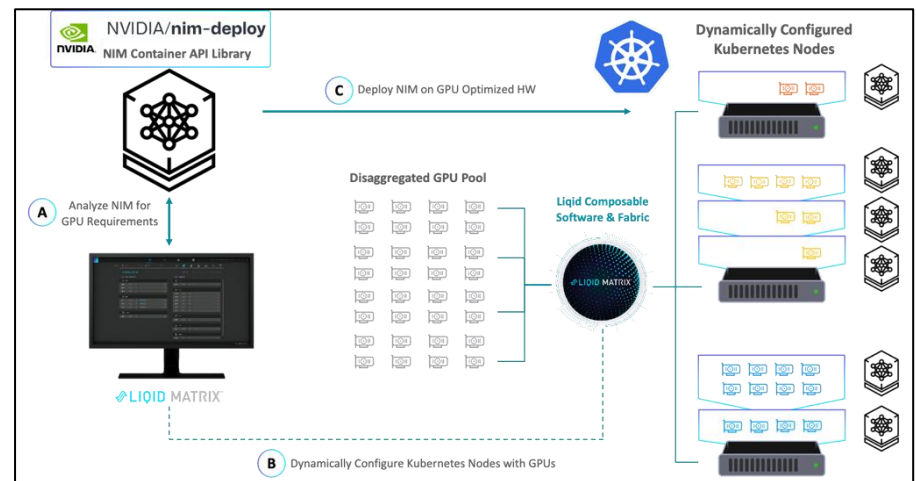
Integración con SLURM y Kubernetes

SLURM Plug-In



- Enable dynamic HW configuration directly from SLURM
- Match HW configuration to workload requirements

Kubernetes Plug-In



Inferencia automática con Kubernetes



kubernetes



openstack.

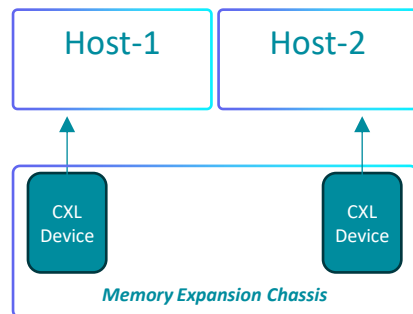


OMNIA

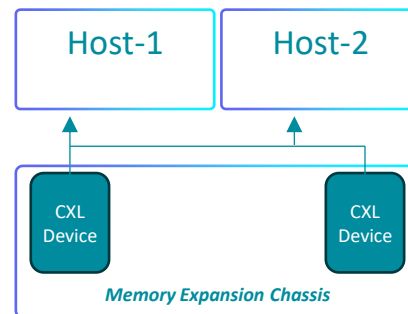
NUTANIX

Liquid también permite expansión de RAM con CxL

Pooling de Memoria

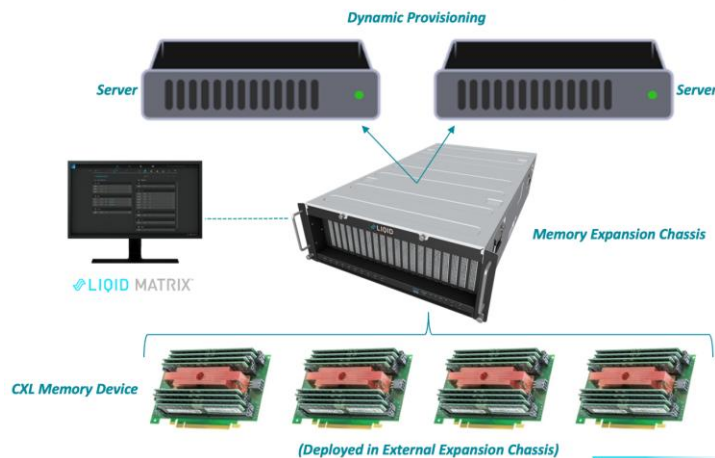


Compartir Memoria



- ✓ Multi-Host Memory Pooling
- ✓ External Expansion Chassis
- ✓ Dynamic Provisioning Per Host

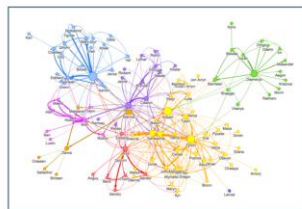
- ✓ Simultaneous Memory Sharing
- ✓ Enable Application Interleaving
- ✓ Performance, Features, TCO



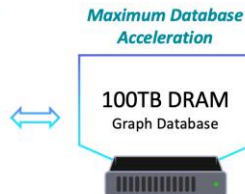
Beneficios de componer RAM dinámicamente

Hasta 100TB RAM/servidor

Graph Analytics and Neural Networks



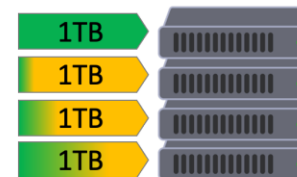
Larger memory keeps more data in-memory, reducing disk access and accelerating queries



- Up to 100x Faster Queries
- No Sharding up to >100TB
- Zero Network Latency

Reduce la cantidad de RAM comprada

Legacy Infrastructure



Low Utilization
Static Configuration

LIQID



High Utilization
Dynamically Configurable

Usa DIMMs más baratas



256GB x 16 = 4TB

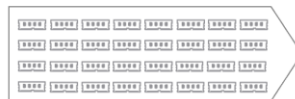


4TB Server – 256GB DIMM

256GB DIMM = \$3,850
16x \$3,850 = \$61,600 / TB
(Expensive Memory DIMM)



64GB x 64 = 4TB

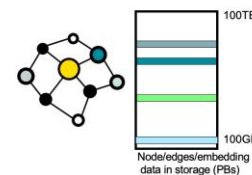


4TB Server – 64GB DIMM

64GB DIMM = \$350
64x \$350 = \$22,400 / TB
(Lower Cost Memory DIMM)

Reduce hasta 25x el tiempo por workload

Graph analytics and neural networks



~8 Hours
NVMe

~20 min
CXL

