

C Columna

Infraestructura, IA y soberanía de datos



Por José Miguel Guzmán.
co-founder Whitestack

Durante 2025, el cloud computing dejó de ser sinónimo exclusivo de nube pública para dar paso a un ecosistema más diverso y matizado. Más que una corrección de rumbo, el año consolidó una reconfiguración profunda del modelo, impulsada por nuevas exigencias tecnológicas, con la inteligencia artificial (IA) como principal catalizador; una revisión crítica de los costos; y la necesidad estratégica de recuperar control sobre los datos.

El resultado fue una nueva edad dorada del datacenter, marcada por el resurgimiento de un modelo on-prem moderno y la consolidación de clouds privados diseñados bajo principios cloud-native.

Las inversiones en infraestructura de centros de datos alcanzaron niveles históricos. Los grandes hyperscalers continuaron expandiendo capacidad, pero con un foco definido: plataformas optimizadas para cargas de IA, mayor densidad de cómputo y arquitecturas concebidas desde su origen para workloads intensivos en GPU y aceleradores especializados.

En paralelo, las empresas de sectores regulados como banca, telecomunicaciones,

energía y sector público, retomaron con fuerza las apuestas por entornos privados, que superaron los US \$100 mil millones y, según estimaciones de mercado, podrían alcanzar los US \$200 mil millones hacia 2030. Ya no se trata del on-prem tradicional, sino de plataformas automatizadas, basadas en software y con interfaces compatibles con la nube pública.

La repatriación de workloads se aceleró por tres factores clave: mayor previsibilidad de costos, soberanía de los datos y menores latencias para nuevos casos de uso.

La IA fue el gran catalizador del año. Modelos más grandes, inferencias más frecuentes y aplicaciones en tiempo real obligaron a replantear dónde y cómo se ejecutan las cargas de trabajo. En este contexto, el edge computing habilitado por el despliegue de redes 5G, comenzó a ganar protagonismo. Industria, ciudades inteligentes, retail avanzado y analítica en tiempo real impulsaron inversiones en micro-datacenters y nodos edge integrados a redes de telecomunicaciones, muchos de ellos bajo esquemas privados o soberanos.

2025 también marcó la consolidación de los llamados Neo

Clouds: proveedores especializados en infraestructura para IA, con arquitecturas altamente optimizadas, clusters de GPU de última generación y esquemas de consumo más transparentes. Este fenómeno debilitó el dominio de facto de algunos hyperscalers y aceleró la innovación tanto en hardware como en modelos operativos.

También el hardware evolucionó de forma decisiva. Procesadores con más cores, mejoras sustanciales en eficiencia energética y la irrupción de nuevos fabricantes de GPUs y aceleradores redujeron la dependencia de pocos actores dominantes, fortaleciendo la competencia y mejorando la sustentabilidad de instalaciones cada vez más densas y exigentes en términos de refrigeración.

La sostenibilidad dejó de ser un discurso para convertirse en una variable de diseño. Energías renovables, enfriamiento líquido, reutilización de calor y optimización mediante software pasaron a formar parte del estándar operativo. En muchos casos, plataformas privadas bien diseñadas demostraron ser más eficientes que despliegues públicos sobredimensionados.