



SOLUZIONI ADOTTATE

- ✓ [Data Center Infrastructure](#)

ARUBA ENTERPRISE PER XREFERENCE OLTRE IL CLOUD: UN'INFRASTRUTTURA SOVRANA PER L'INFERENZA AI

Xference rivoluziona l'interazione con la tecnologia attraverso l'inferenza AI, proponendosi come Inference Provider focalizzato sull'assoluta privacy e sulla piena sovranità sui dati. Fondata nel 2025 da Andrea Pili, Michele Fadda e Francesca Audino, la startup risponde a una sfida critica di importanza crescente: consentire ad aziende di ogni dimensione, enti e professionisti di sfruttare la AI senza esporre all'esterno dati sensibili e proprietari. La soluzione è un'infrastruttura costruita intorno alla conformità alle normative rilevanti, come GDPR e AI Act, e coniuga prestazioni avanzate, sicurezza e accessibilità per rendere la AI locale e su misura disponibile a realtà di ogni dimensione in modo rapido, facile ed economico. Il progetto Xference ha sede a Cagliari, città inserita nel solco di una ricca tradizione locale di innovazione nel settore delle telecomunicazioni e delle reti dati. La società ha concluso con successo un round pre-seed di 700.000 euro che ha permesso l'avvio della fase beta a inizio 2026 con i primi 100 clienti su invito, e prevede di completare entro la fine dello stesso anno un round seed per alimentare la crescita controllata dell'infrastruttura, del servizio e del business.

La necessità

Xference è una ambiziosa startup nata con l'obiettivo di risolvere il paradosso della AI contemporanea: il ricorso a infrastrutture globali, opache ed energivore per ogni necessità di inferenza, anche la più semplice e mirata. Un approccio inefficiente che comporta un sovradimensionamento del perimetro informativo, dipendenza da connessioni globali e rischi per la privacy a fronte di costi e consumi inutilmente eccessivi. "La nostra idea è invece quella di fornire un'infrastruttura per l'inferenza di modelli LLM open source che garantisca la completa autonomia digitale e la massima privacy", afferma Andrea Pili, cofondatore e CEO di Xference. "A differenza dei provider tradizionali, infatti, non ci occupiamo di addestrare modelli ma di ottimizzarne l'esecuzione, assicurando che i dati dei clienti restino confinati nei rispettivi ambienti senza dover essere mai duplicati o trasferiti esternamente".

L'obiettivo è facilitare l'adozione della AI tramite soluzioni no-code, efficienti dal punto di vista energetico e conformi alle normative

europee come GDPR e AI Act. Installando LLM open source su macchine dedicate collegate tramite una connessione criptata end-to-end e site-to-site, il processo di inferenza rimane sotto il controllo esclusivo del cliente senza che nessuno possa accedervi esternamente.

Escluso il ricorso al cloud per inadeguati livelli di trasparenza e controllo, Xference ha scelto anche di evitare la costruzione di siti fisici proprietari preferendo invece affiancarsi a un partner esterno che possa fornire un data center sicuro, scalabile e performante, adeguato alle necessità di un'infrastruttura in crescita. "In particolare abbiamo bisogno di garantire tutta la potenza necessaria ai nostri sistemi – parliamo di un fabbisogno di 24kW per rack quando l'offerta tipica nel settore è pari a circa un quarto – unitamente a una scalabilità lineare che ci consenta di passare dalle dimensioni limitate di startup a Inference Provider europeo di riferimento con centinaia di nodi attivi senza incontrare problemi di saturazione dello spazio fisico", ha sottolineato Pili.

La soluzione

Aruba ha rapidamente disegnato e dispiegato una soluzione di colocation capace di sostenere la traiettoria di crescita di Xference senza introdurre limitazioni strutturali di alcun genere, aggiungendovi una proposta Hardware-as-a-Service (HaaS) che si fa carico dell'approvvigionamento e dell'attivazione operativa delle macchine secondo le specifiche configurazioni richieste dal cliente. "Questo aspetto è molto importante specie in fase di startup perché ci evita di dover seguire

attività che, pur essenziali, rappresentano una distrazione rispetto al nostro focus principale", ha rilevato Michele Fadda, fondatore e CTO. "Aruba è riuscita a coprire anche questa esigenza, pur non essendo semplice approvvigionare sistemi secondo le nostre specifiche. La risposta che abbiamo ottenuto complessivamente è quella che ci ha fatto capire di aver trovato non un semplice fornitore, bensì un vero partner capace di accompagnarci nella nostra avventura".

I vantaggi

La partnership si traduce in una serie di vantaggi strategici, tecnici e finanziari fondamentali per una startup che vuole accelerare il proprio sviluppo. L'insediamento nel data center IT3 di Ponte San Pietro colloca l'infrastruttura di Xference in un contesto di eccellenza dalle ampie disponibilità di spazio e potenza che garantirà la crescita mantenendo fisicamente concentrati gli apparati hardware, così da ridurre latenze e complessità di gestione. Gli stessi apparati sono erogati in modalità HaaS da Aruba, accorgimento che consente di evitare investimenti in conto capitale trasformandoli in più agili costi operativi; così come è la stessa Aruba a occuparsi di tutti gli aspetti dell'installazione

fisica, della configurazione dei rack e della gestione delle componenti di networking (firewall perimetrali e switch).

"La proposta di Xference mira a rendere il cliente pienamente sovrano e indipendente nell'utilizzo della AI", spiega Andrea Pili. "Tuttavia la sovranità digitale non può essere ridotta al semplice fattore geografico stabilendo un data center in Italia, ma deve garantire l'autonomia e la riservatezza totale del dato. Questo obiettivo è pienamente condiviso da Aruba, una eccellenza italiana controllata da una proprietà nazionale che può vantare numerose certificazioni orientate alla sicurezza, alla privacy e alla conformità a ogni livello".

Sviluppi futuri

Una volta terminata la fase preliminare sui primi 100 utenti beta, Xference avvierà un lancio controllato a inviti per crescere gradualmente monitorando i carichi di lavoro in tempo reale, così da ottimizzare l'ampliamento delle risorse in base alle effettive necessità dell'utenza. Una volta a regime, gli apparati hardware potranno essere distribuiti per creare un'infrastruttura

edge in grado di posizionare i nodi di calcolo il più vicino possibile all'utente finale allo scopo di ridurre ulteriormente le latenze. Per i clienti enterprise con requisiti di controllo dei dati ancora più stringenti sarà infine messa a disposizione una speciale modalità housing riservando a uso esclusivo specifiche macchine del data center.

