

Healthcare — HIPAA Radiology Inference

Client: US Hospital Network · **Region:** NA

Workload

3D UNet + report generation LLM (13B)

Deployment topology

34 Hotel Edge + 2 Core IDC (Ashburn, Dallas)

Measured results

Metric	Value	Note
Report turnaround	-58%	median
PHI egress	0	bytes
Sites onboarded	34	in 6 months

Methodology

Figures anonymised at customer request. Latencies measured at the client SDK boundary over rolling 14-day windows. Cost deltas reference the customer's prior production stack, normalised per token-equivalent.