SOLARVOLT

WHITEPAPER

Cloud or Colo? The point when African AI startups need to make the switch.

Why the economics of AI compute break on the public cloud at scale — and why, for African startups, they break sooner and harder.

A SolarVolt whitepaper · SolarVolt builds and operates the AI data centre within Mawuko Technology Park, developed by site owner Integrated Uwa Services Limited (IUS)

EXECUTIVE SUMMARY

It is far more economical for AI startups running at scale to move most or all of their workloads from the public cloud to a colocation provider. As a rule of thumb, **50% continuous utilisation is the cut-over point** at which owning (or renting) your own iron in a colo beats renting capacity from a hyperscaler.

That is true everywhere. For an **African** AI startup it is truer still. You earn in Naira, Cedi or Shilling but pay your cloud bill in US dollars; you move large training datasets across oceans and pay punishing egress fees to get them back; you serve local users from a data-centre region thousands of kilometres away; and, increasingly, data-protection law requires that some of your data never leaves the country at all. Every one of those forces pulls the cloud-to-colo switch **forward**.

> The barrier to African colocation has never really been demand, or talent, or even capital. It has been power. SolarVolt removes it — by contracting the data centre's power for it, and building the park around generation rather than around the grid.

01 The easy on-ramp — and the illusion

Any startup can begin on a public cloud with a credit card, a dream and an algorithm to test. Spin up GPU-accelerated instances, train a model, tear it down. AWS, Azure, Google, Alibaba and the rest have made this genuinely easy — and easier still to grow into what looks, to a young company, like always-available, infinite capacity.

It is an illusion the cloud giants work hard to maintain. There is no such thing as perfectly fungible, infinite capacity — and no AI startup was ever denied funding for running on a major cloud. But once the prototype works and the bills start arriving, two realities sink in fast: **data gravity** (training sets grow into the terabytes and petabytes) and **data egress** (moving that data *off* a cloud is slow and astonishingly expensive). Pick one provider's tooling, compute, storage and networking, and you have the same vendor lock-in that a proprietary minicomputer represented in the 1990s.

02 The African cost problem — sharper on every axis

The generic cloud-vs-colo argument understates the case for an African startup. Four multipliers stack on top of it:

- **Foreign-exchange exposure.** Cloud is invoiced in USD. When your revenue is in a local currency that is depreciating, your effective compute cost rises even when your usage is flat. Owning capacity fixes cost in a way a dollar bill never can.
- **Egress and data gravity across oceans.** Your data lives where your users are. Round-tripping training and inference data to US or EU regions means both latency and egress fees on every gigabyte pulled back.

- **Latency to distant regions.** Serving real-time inference (a fraud check, a clinical-AI result, a voice model) from a region thousands of kilometres away adds tens of milliseconds that in-country compute simply removes.
- **Data sovereignty.** Nigeria's Data Protection Act 2023 and comparable regimes across the continent are pushing regulated workloads onshore. For health, financial and government data, "just use a US region" is no longer a compliant answer.

> For an African AI company at production scale, the question isn't *whether* to move off the cloud. It's whether there is anywhere onshore, with reliable power and Tier III uptime, to move to.

03 Doing the maths: colo vs cloud for AI

GPU-accelerated capacity is expensive to rent. Worse, a web-scale cloud is virtual compute, storage and networking scattered across a 100,000-server region — a sub-optimal platform for AI workloads that need deterministic, genuine performance. It is telling that NVIDIA runs its Selene supercomputer in its own data centre, and that Google does not run its core machine-learning workloads on Google Cloud's public Kubernetes layer. The people who *built* the clouds keep their heaviest AI iron on dedicated infrastructure.

In a multi-tenant cloud, demand often exceeds supply and one tenant's spike becomes your slowdown — "noisy neighbours." The risk of performance degradation grows exactly as a startup climbs the hockey-stick curve it hoped to ride. When an AI workload is in production and driving revenue, **predictable** performance and **better** performance both become non-negotiable — and both are hard to get on the cloud, at a reasonable price, all the time.

The 50% rule and the hybrid model

Before renting a single GPU instance, map how your compute needs evolve from prototype → product → production. Capacity grows over time on any metric you choose — MFLOPs, epochs, or kilowatts. Anytime you are using cloud capacity at scale for more than about **50%** of any appreciable long term — a year or two or three, even with reserved instances — it is far more economical to run a **base production load on iron you own or lease in a**

colocation facility, and burst into multiple clouds only for peaks. You buy the time to grow on-premises capacity by bursting to cloud, without tying yourself to any single vendor or plan.

04 TCO: an illustrative comparison

Take an eight-GPU AI server (e.g. an NVIDIA DGX-class node) as the unit of comparison. In a colocation facility, the hardware plus storage and switching — depreciated straight-line over two years — lands near `$10,000/month`, plus roughly `$2,000/month` for ~10 kW of power and rack space: about **$12,000/month all-in**. The equivalent on-demand cloud instance can approach `$1,000,000 over two years` once storage is included.

CLOUD UTILISATION	ON-DEMAND CLOUD (2-YR)	1-YR RESERVED CLOUD	COLO / IN-HOUSE
25%	$250,000	$586,200	$288,000
50%	$500,000	$586,200	$288,000
80%	$800,000	$586,200	$288,000
100%	$1,000,000	$586,200	$288,000

Illustrative framework (2-year, ~$12k/month colo baseline), adapted from public cloud vs colocation comparisons. Figures are directional, not a budget — but the conclusion is inescapable: past ~50% utilisation, colocation wins, and it widens from there. For an African startup, add the FX and egress multipliers of Section 02 on top of the cloud columns.

05 The power problem — and why it has stopped African colo before

Everything above assumes a colocation facility exists that can actually keep the racks running. In much of Africa, it hasn't — because grid power is unreliable, and running a Tier III AI facility on diesel alone is neither economic nor sustainable. That single constraint has kept the continent's AI workloads stranded on distant clouds.

The SolarVolt data centre is built to remove it at the root — because the park it sits in was designed around power. **The data centre's default supply is an on-site solar and battery microgrid under a licensed off-grid PPA**, with SolarVolt-owned N+1 generators able to carry the full load. The park's compounds are laid out for up to 100 MW of co-located generation as capacity scales, and IUS is in discussion with Tetracore Energy — whose Atakobo Independent Power System in Ogun State began generating in April 2026 — on gas-fired supply. Tenants of SolarVolt's data centre are insulated from grid instability by design — power is contracted and generated for the park, not drawn from a fragile national grid.

Up to 100 MW	Tier III	130 kW	<1.4
Compound capacity for co-located generation	Concurrently maintainable	Per-rack ultra-high AI density	Target PUE

06 The SolarVolt campus

SolarVolt is building and operating a 20 MW Tier III, AI-ready data centre inside **Mawuko Technology Park**, Odeda LGA, Ogun State — a park developed by site owner **Integrated Uwa Services Limited (IUS)**, which provides the land, regulatory compliance, power and amenities. The data centre is phased 5 MW + 5 MW + 10 MW and engineered from day one for GPU-accelerated AI and high-performance computing.

- **Power engineered for the load:** on-site solar + battery microgrid under a licensed off-grid PPA, with SolarVolt-owned N+1 generators standing by; park compounds laid out for up to 100 MW of co-located gas generation as capacity scales.
- **Ultra-high density:** halls engineered up to 130 kW per rack, with direct-to-chip liquid-cooling readiness — a class above standard colocation.
- **Air + liquid cooling** for AI workloads, targeting a PUE under 1.4 in a warm climate.
- **Carrier-neutral, dual-fibre entry** — you choose your own networks and cloud on-ramps, and burst to the public cloud only when it pays to.

- **Sovereign by location:** in-country compute that keeps regulated data onshore under the NDPA 2023, with the latency of local hosting.
- **GPU-as-a-Service or dedicated bare metal** — start with consumable AI compute, grow into your own single-tenant hall.

It is the combination African AI has been missing: hyperscale-grade density and uptime, backed by power the operator owns rather than hopes for.

MAKE THE SWITCH

Reserve capacity in Nigeria's AI data centre.

Whether you need a few GPUs, a rack, or a dedicated AI hall, the maths — and the map — point the same way. Talk to us early; Phase 1 capacity is being allocated to anchor tenants now.

Talk to our team

↓ Download this whitepaper (PDF)

About SolarVolt & IUS. SolarVolt is building and operating a 20 MW Tier III, AI-ready data centre and securing its tenants. The data centre sits within Mawuko Technology Park, Ogun State, Nigeria, developed by **Integrated Uwa Services Limited (IUS, RC 8282388)** — the site owner responsible for land acquisition, regulatory compliance, and park amenities including the power compounds, laid out for up to 100 MW of co-located generation, and water.

Contact: Andrew Okunzuwa, Business Development · andrewokunzuwa@solar-volt.co.uk · +44 7748 431133

This whitepaper adapts the industry "cloud vs colocation" thesis for the African AI market. Cost figures are illustrative frameworks, not quotations.