

FRAMEWORK #132

Coordination Density

The CPU is back in the stack — and nobody budgeted for it.

AI INFRASTRUCTURE

THE INFRASTRUCTURE RATIO HAS REVERSED

GPUs execute. CPUs coordinate.

TRAINING / BATCH ERA

CPU COORDINATION — 15%



GPU EXECUTION — 85%

AGENTIC ERA

CPU COORDINATION — 65%



GPU EXECUTION — 35%

Coordination Density

The amount of orchestration, governance, retrieval, and policy evaluation work required to produce a unit of AI execution.

Compute Density = how much infrastructure executes work.

Coordination Density = how much infrastructure is required to govern that work.

What coordination density actually measures

01

Orchestration

Agent coordination and call sequencing

02

Context Assembly

Retrieval, memory, and tool state

03

Policy Evaluation

Authorization and governance checks

04

Governance Enforcement

Allow, deny, redirect, or escalate

THE LAYER SPLIT

**GPUs execute intelligence.
CPUs govern intelligence.**

AI EXECUTION

GPU

AI GOVERNANCE

CPU

The industry is measuring the wrong resource

TRACKED

- GPU utilization
- GPU memory / VRAM
- Network fabric

UNTRACKED

- Orchestration call volume
- Policy evaluation latency
- Context assembly overhead
- Host CPU saturation

Capacity planning has to catch up

- Host CPU sizing follows agent fan-out and policy chain depth — not GPU attach ratio
- Policy and orchestration layers get their own capacity headroom
- Governance latency becomes a first-class SLO
- Procurement risk shifts toward host-side compute as agentic adoption grows

FRAMEWORK #132

The next constraint won't look like the last one.

*Compute density was the first generation's constraint.
Coordination density is the next one.*

Rack2Cloud.com | Follow [The Architect](#) for more