LangChain Agent Write-Path Benchmark

ana ABS vs AWS EBS (gp3 & io2) · state-serialization checkpoint writes

measured the read path. Part 2 measures the write path: identical LangChain checkpoint workloads on the same orage platforms, where every agent step blocks on fsync before it can proceed.

S LEAD

24-28%

ster end-to-end vs
est AWS at sustained
ad

THE WORKLOAD

Checkpoint commits

Postgres · Redis · Qdrant

WAL fsync · AOF append · vector upsert (wait=true)

3 disk-bound commits per task

REPRODUCIBILITY

18 runs, 18 wins

CONCURRENCY

1,000 agents × **100** tasks

100,000 checkpoints total

×

er commit p99

gres commit p99 at 100k checkpoints: 105-107
s 210-232 ms.

33-34%

Lower task p99

918-925 ms vs 1,332 ms+ (write) at sustained load.

31×

Cheaper than io2

$118/mo vs $3,710/mo (io2-64k).

HE SETUP · SAME FIVE PLATFORMS AS PART 1

IG	INSTANCE	VCPU / RAM	STORAGE	PROVISIONED IOPS	TOTAL COST
vana ABS	n1-standard-4	4 / 16 GB · DDR5	ABS 256 GB	20,000 baseline; 600,000 burst, included	$
-3k	m6i.xlarge	4 / 16 GB · DDR4	gp3 256 GB	3,000	$
-16k	m6i.xlarge	4 / 16 GB · DDR4	gp3 256 GB	16,000	$
-32k	m6i.xlarge	4 / 16 GB · DDR4	io2 256 GB	32,000	$2,
-64k	m6i.xlarge	4 / 16 GB · DDR4	io2 256 GB	64,000	$3,

ch task is one checkpoint cycle: Postgres INSERT + COMMIT (WAL fsync), Redis SETEX (AOF), Qdrant upsert of 5 vectors with `wait=true`.
load 2 KB + 2 KB per step, capped at 64 KB.

o modes, back-to-back: `write` is pure checkpoint cycles; `mixed` reads context first, then commits. 10 concurrent workers, 5M preloaded ctors across 50 collections, same as Part 1.

rability is pinned, not tuned: `synchronous_commit=on`, AOF persistence, `wait=true` upserts on all five. No platform gets a durability count.

NIRVANA LABS

RESULTS // 100,000 CHECKPOINTS × 2 MODES × 3

ORKLOAD DURATION · WRITE MODE, STEADY STATE · ABS WINS EVERY SCALE

LE	CHECKPOINTS	NIRVANA ABS	IO2-32K	IO2-64K	GP3-16K	GP3-3K	ABS
×10	1,000	36s	39s	40s	40s	40s	
×20	10,000	364s	396s	415s	406s	401s	
0×100	100,000	75 min	105 min	111 min	107 min	102 min	2(

T 100K CHECKPOINTS · BOTH MODES · DURATION AND TASK P99

FORM	WRITE DURATION	WRITE TASK P99	MIXED DURATION	MIXED TAS
ana ABS	75 min	919 ms	84 min	9
32k	105 min	1,392 ms	112 min	1,4
64k	111 min	1,449 ms	120 min	1,5
-16k	107 min	1,389 ms	114 min	1,5
-3k	102 min	1,392 ms	147 min	2,6

state run (R3). **gp3-3k is the only platform where mixed is slower than pure write** — the added read IOPS push it past its 3,000-IOPS ceiling.

ER-SERVICE COMMIT P99 · 1000×100 SUSTAINED · STEADY STATE

IT PATH	ABS P99	IO2-32K P99	IO2-64K P99	GP3-3K P99	ABS \
gres commit · write	105-107 ms	210-213 ms	224-232 ms	217-219 ms	2×
nt upsert · write	254-282 ms	371-381 ms	395-398 ms	667-699 ms	32%
gres commit · mixed	79-82 ms	158-163 ms	170-171 ms	166-175 ms	2×
st 10s-window p99	429-536 ms	600-678 ms	620-744 ms	566-664 ms	Lowest abs(

heckpoints io2-32k and gp3-16k still hold the best per-op commit tail (78-88 ms vs ABS 95-105 ms). **At 100k it inverts: ABS holds half the commit p99 and the lowest absolu indow.**

HE VERDICT · THE WRITE PATH

STER

24-28%

ster end-to-end at
ustained load, with every
ommit waiting on disk.

UNDER LOAD

First in **all 18 runs** — write and mixed, every scale. Margin grows from **~10%** to **24-28%** as load climbs.

PRICE

$118/mo

31× cheaper than $3,710 (io2-64k).

PEN SOURCE · RUN IT YOURSELF

b.com/nirvana-labs-examples/langchain-benchmarks · state-serialization-writes

ry step commits.
r agents. Read and write faster.

IRVANA LABS

@nirvanalabs_bd nirvanalabs.io sales@nirvanalabs.io @nirvana