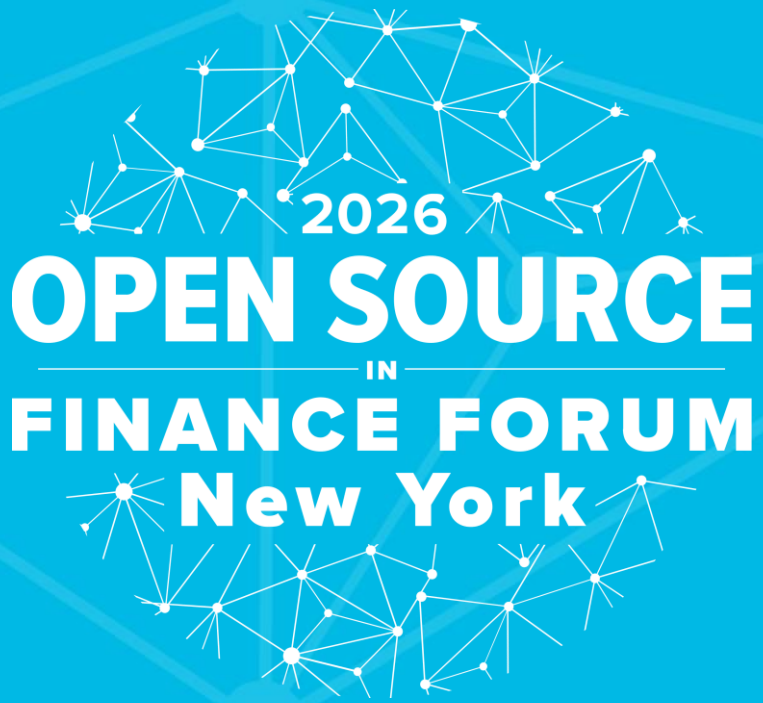


Benchmarking Fluxnova for Financial and Agentic Payment Workflows

Repeatable tests for performance, reliability and correct outcomes

Nicholas Cusato
Oracle Cloud Infrastructure

Introduction

Nicholas Cusato
Senior Cloud Engineer

Oracle Cloud Infrastructure

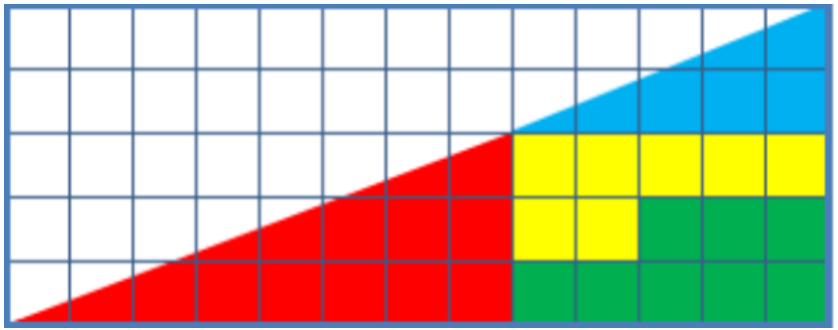
Financial Services and Insurance

Cayucos, CA (central coast)

Research, Sales, DevRel
Customer Onboarding / Enablement
Benchmarking – Price & Performance



A Modernization Journey



What should we keep?

Where could AI help?

How will we know it works better?

Trekky0623 at English Wikipedia, Public domain, via Wikimedia Commons

Speed is only part of the answer

"A fast agent that selects an unauthorized tool, **duplicates a payment** or reaches the wrong final state **has failed.**"

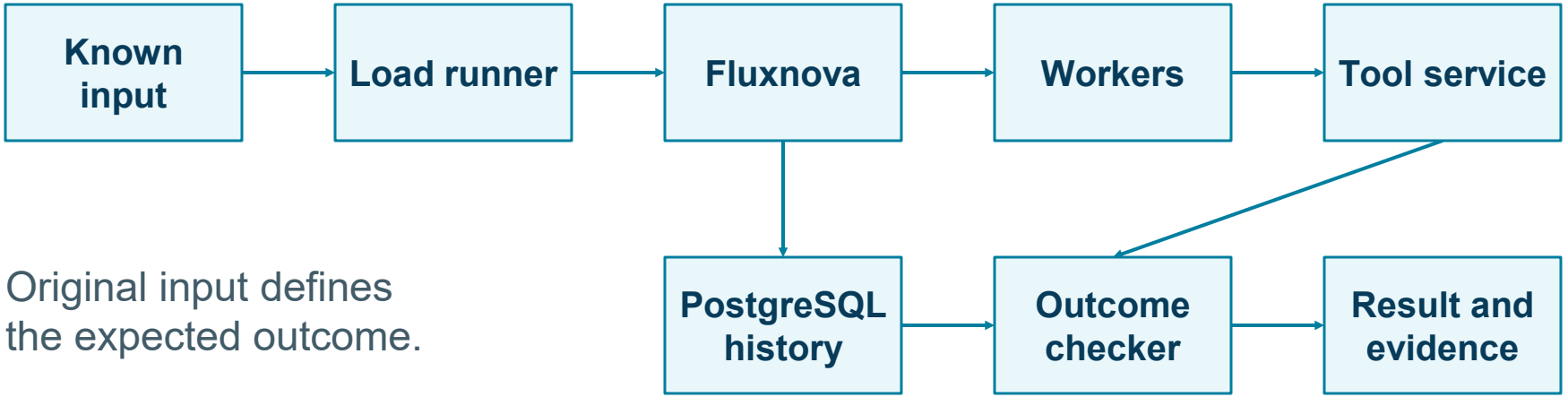
Make workflow changes with evidence

Use repeatable tests to see whether a workflow change improves performance while preserving correct outcomes.



How the benchmark works

The flow runs in Fluxnova. The benchmark checks the outcome.



Original input defines the expected outcome.

Hold other conditions steady. Change one factor. Compare outcomes.

Representative path. Each study records its components and deployment.

Benchmark research map

Research

FLUXNOVA

Confidence in workflow changes

6 areas

- **Workflow design**
- **Runtime upgrades**
- **Load and endurance**
- **Events and recovery**
- **Agent controls**
- **Measurement**

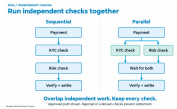
17 study records

CORE	Basic flows
W01	Worker pacing
W02	Parallel checks
W02 mixed	Mixed outcomes
W03	Review routing
PAY05	FX quotes
JVM	Java 21 / 25
RELEASE	Fluxnova releases
P01	Reference baseline
SOAK	Eight-hour run
CAPACITY	Capacity (partial)
FNV06	Events / batches
FNV07	Retries / faults
FNV07 timing	Timing (local)
RECOVERY	State recovery
AGENT	Authority checks
OBSERVER	Polling / timing

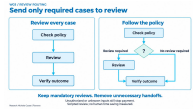
9 workflow diagrams



D01 Worker pacing



D02 Parallel checks



D03 Review routing



D04 Minimal flow



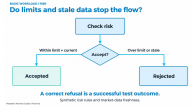
D05 Data / branching



D06 External work



D07 KYC / review



D08 Risk / freshness



D09 Settle / replay

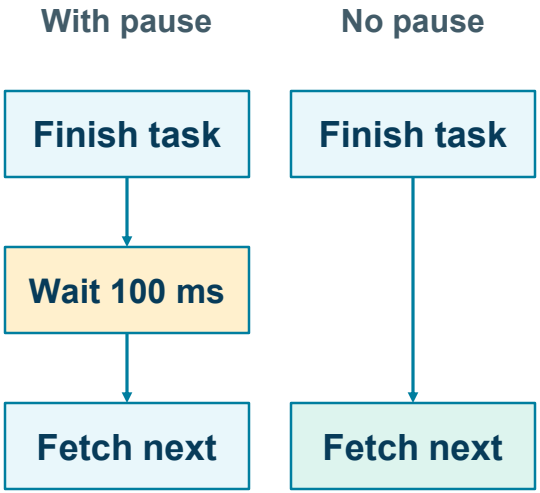
Explore the full catalogue

github.com/oci-fsi-pursuits/fluxnova-benchmark-kit

Completed, partial and diagnostic evidence

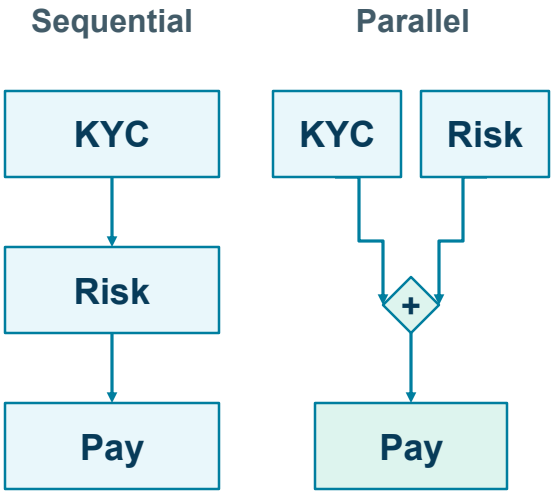
Three workflow comparisons

W01 Worker pacing



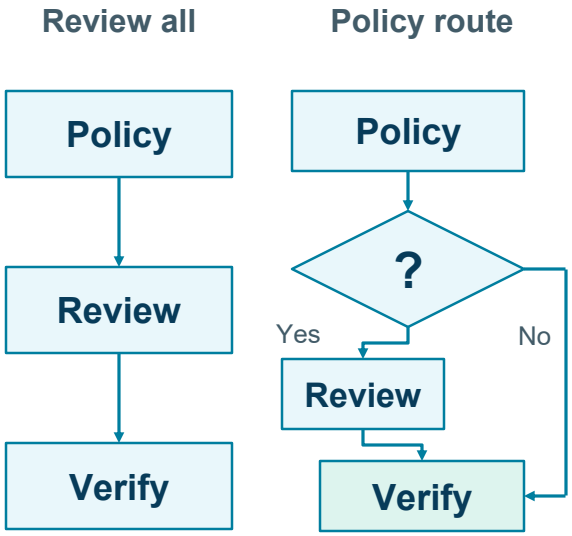
Fetch again after success.
Idle and retry backoff stays.

W02 Parallel checks



Overlap independent checks.
Both finish before payment.

W03 Review routing



Keep required reviews.
Skip unnecessary handoffs.

Synthetic cases. Simplified views of the compared mechanisms.

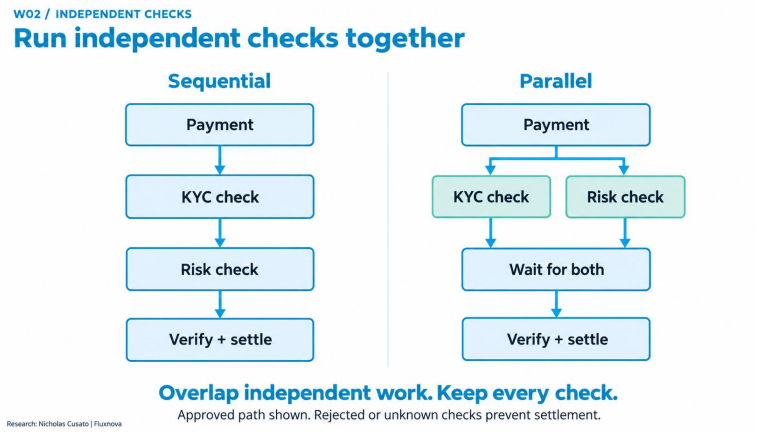
What these tests help you decide.

Area	What we learned
Workflow design	W02 observed p95 fell 32.1% at 1/s. Mixed traffic helped settlements and hurt timely refusals.
Review routing	Review tasks fell from 10 to 4. All four mandatory reviews stayed; reviews were scripted.
Runtime / release	Java and Fluxnova upgrades showed small, mixed differences. No broad speedup emerged on the fixed workload.
Load / endurance	Eight hours: 288,000 correct outcomes within two seconds. Maximum capacity remains unqualified.
Events / recovery	45 event/batch cells completed, with two resource flags. Retry and recovery findings cover specific tested cases.
Agent controls	Functional checks exercised payment authority. A measured live-agent performance benefit remains open.
Measurement	Polling changes when completion is detected. Hold the observation method fixed when comparing.

Each result belongs to its declared workload and environment.

One recipe, one scorecard

The recipe



Pattern: what runs
Benchmark type: the question
Study: one fixed comparison

The scorecard

Field	W02 example
Baseline	Sequential checks
Change	Parallel checks
Keep fixed	Cases, load, workers and observer
Check	Latency and correct outcomes
Explain	Result, environment and limits

Protocol 0.2 is a community proposal. It is not an adopted FINOS standard.

The next research questions

01 Customer workflows and live agents

Representative traffic, real dependencies and an agent-versus-rules comparison.

02 Capacity and replica scaling

Complete load confirmation and qualify the resources used by each run.

03 Platforms and databases

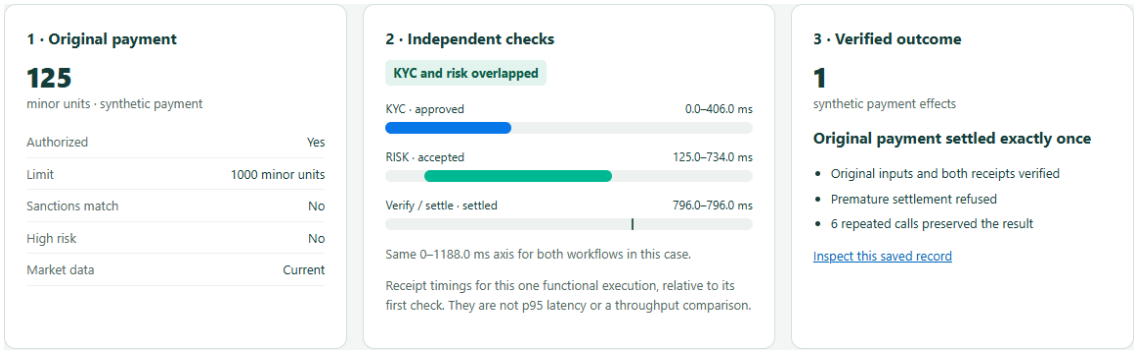
Validate other deployment adapters and compare declared alternatives.

04 Cost and independent reproduction

Allocate the full cost per correct outcome and reproduce a study elsewhere.

Getting Started:

Demo



python -m benchlib.cli explain W02

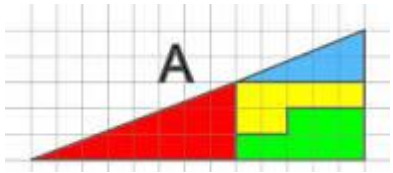
Recorded functional execution. Local setup and OCI Terraform are documented.

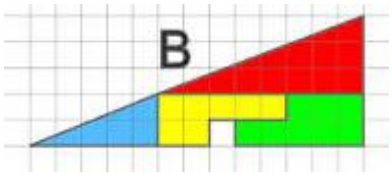
Source and run guide

github.com/oci-fsi-pursuits/fluxnova-benchmark-kit

“Measure twice, Pay once”

Confidence comes from a comparison you can repeat.





Bring your workflow.
Keep its required controls.
Share what changed.

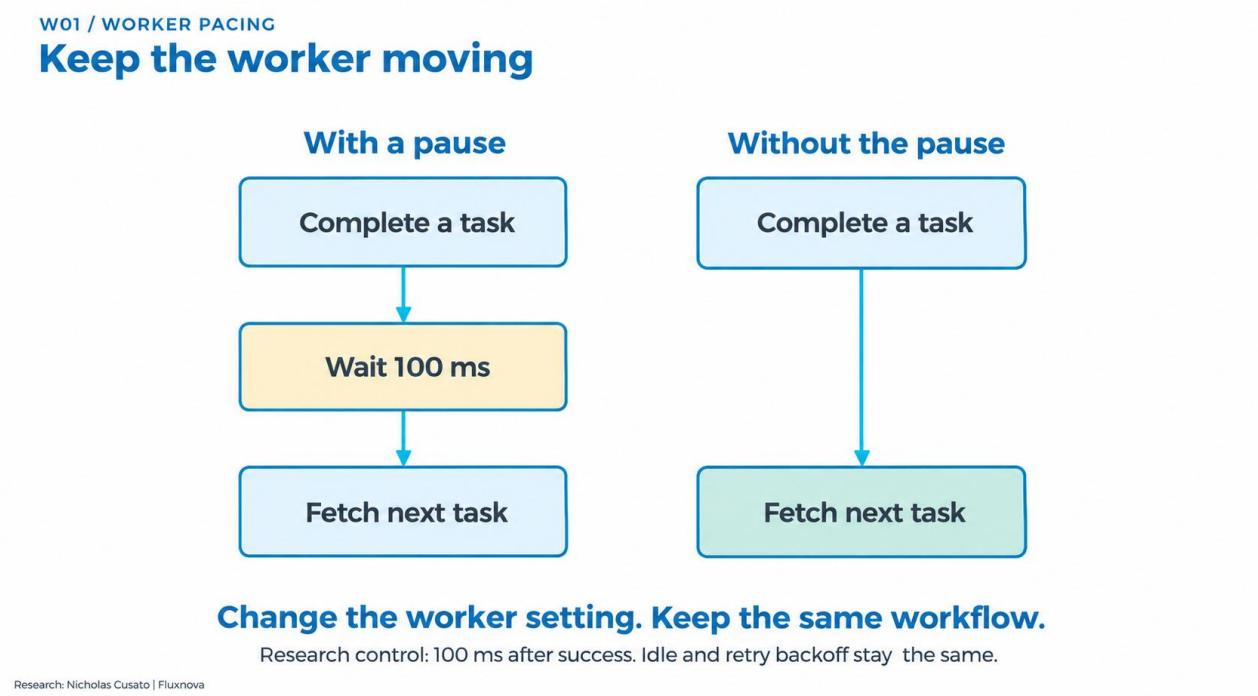
Research and OCI implementation: Nicholas Cusato
Fluxnova software: FINOS contributors

Feedback and next workloads

Benchmarking Fluxnova for Financial and Agentic Payment Workflows

GitHub Issues
Benchmark feedback

D01 · Worker pacing

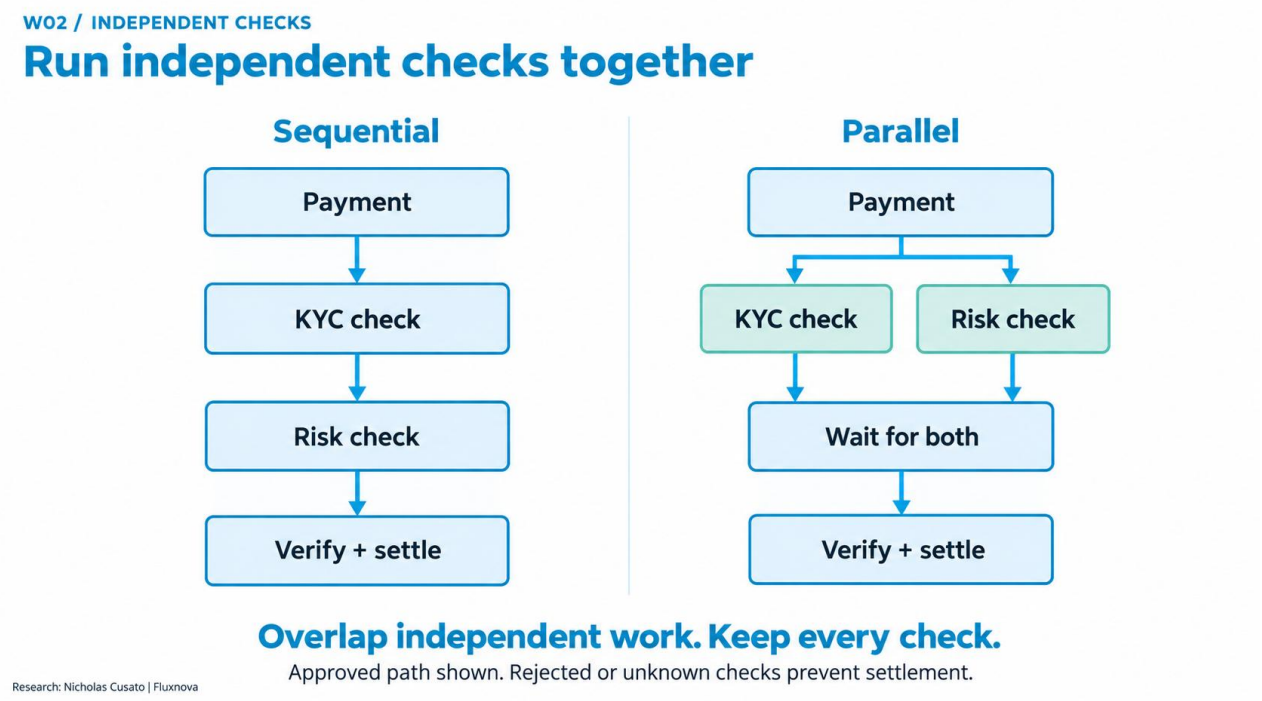


What runs

The payment graph stays the same. The external worker either sleeps after success or immediately fetches the next task. Idle and retry backoff remain.

W01 shows more payments meeting the deadline for fixed bursts and one worker. It tests worker scheduling, not a redesigned payment process.

D02 · Parallel checks

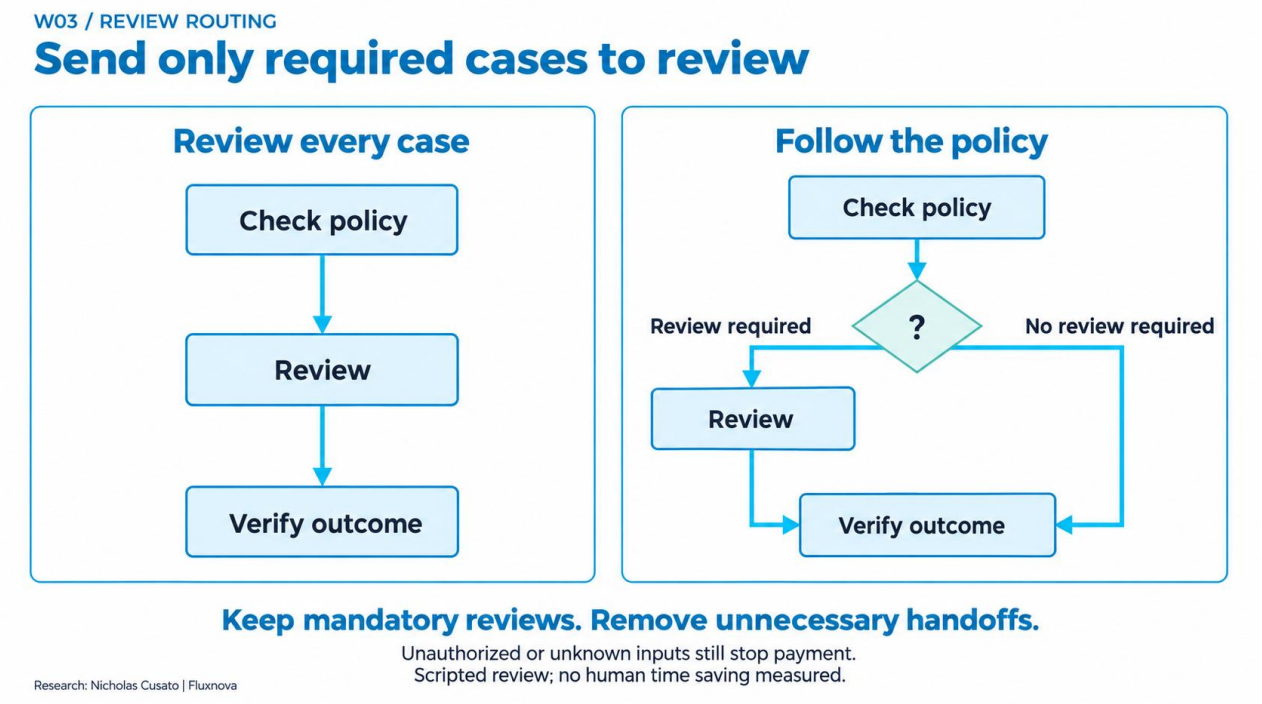


What runs

Sequential execution waits for KYC before starting risk. Parallel execution starts both independent checks and joins them before payment. Neither check is skipped.

W02 shows lower observed latency at 1/s. The mixed follow-up shows that benefits depend on the outcome and dependency delays.

D03 · Review routing

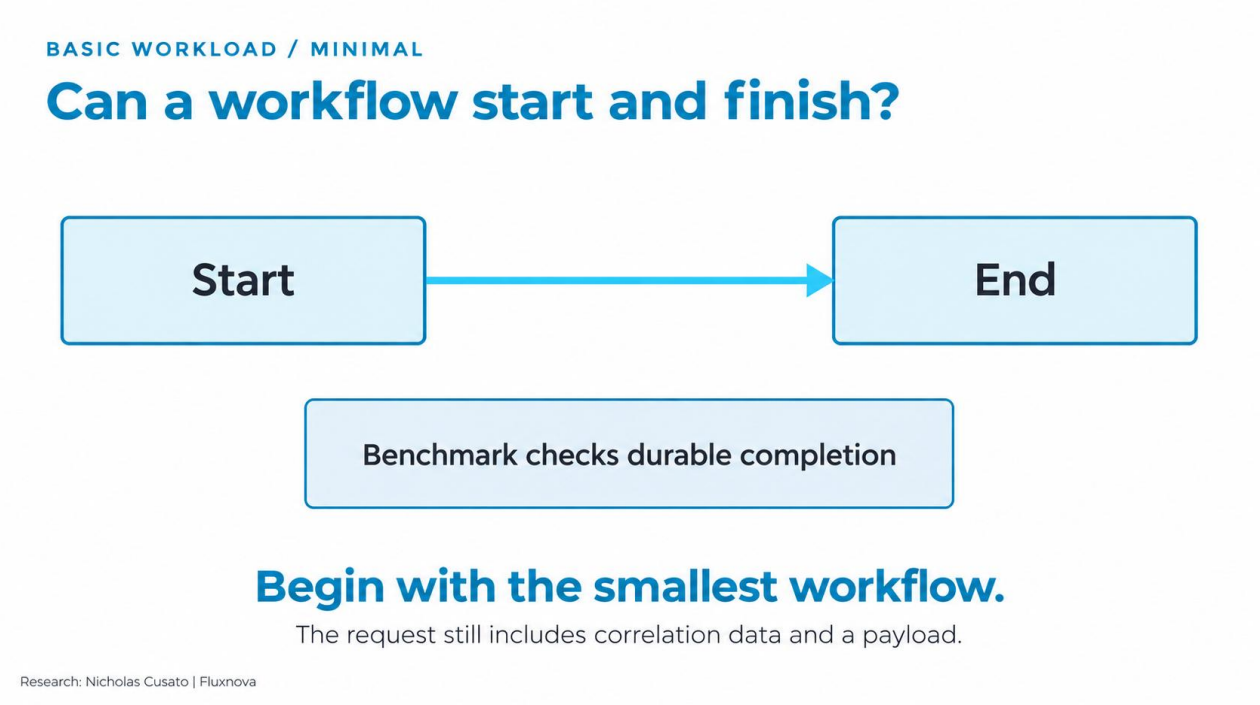


What runs

A fixed policy routes required cases to a reviewer. Eligible cases can proceed; unauthorized or unknown cases still stop. Final verification remains.

W03 reduces tasks from ten to four for its constructed cases and preserves all four mandatory reviews. It does not measure human time saved.

D04 · Minimal workflow

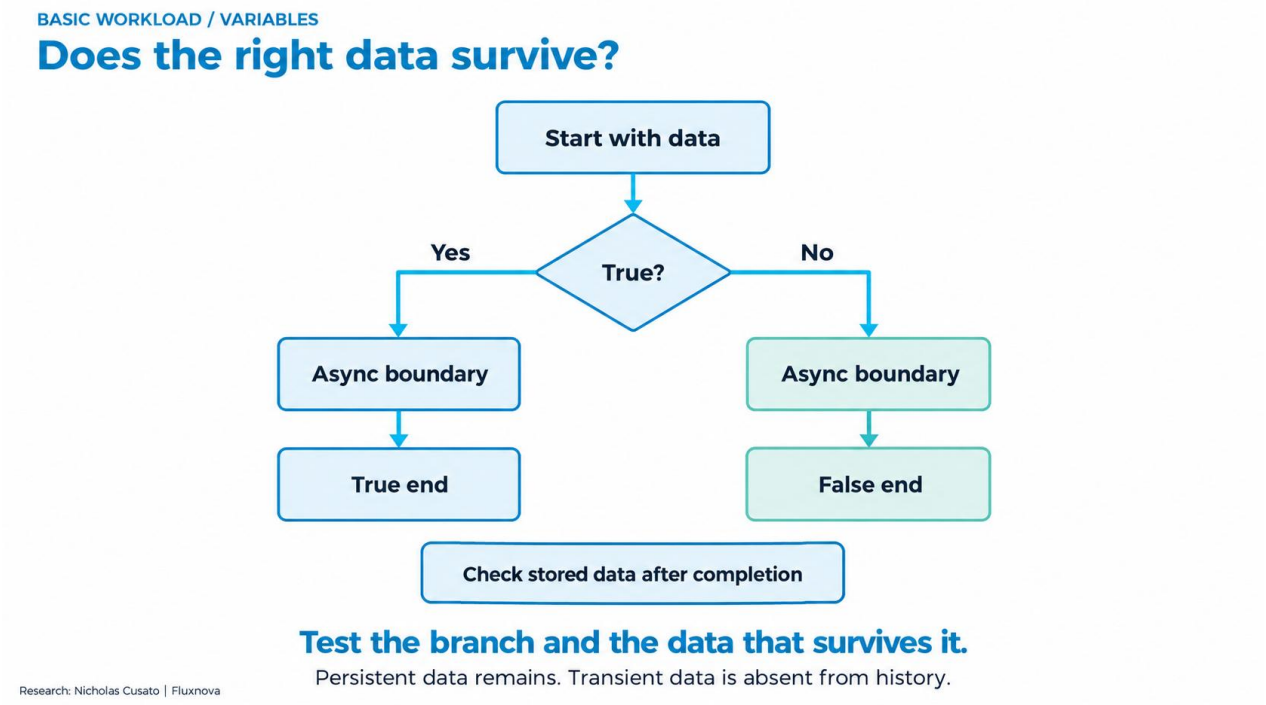


What runs

Start and finish a process with the declared correlation data and payload. The checker confirms durable completion.

A useful basic readiness check. Its input still contains data, so it is not a zero-variable engine-overhead benchmark.

D05 · Variables and branching

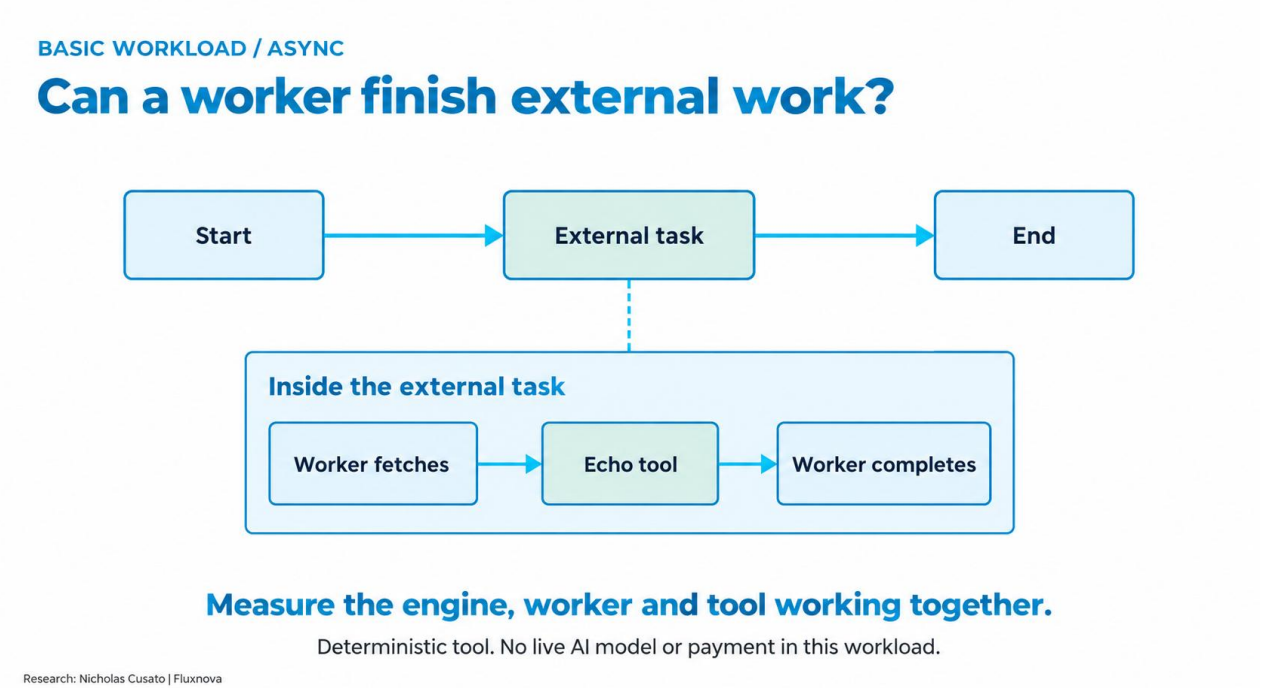


What runs

Choose a branch, cross an asynchronous boundary and inspect persisted values. Persistent and transient data have different expected histories.

Correctness requires the right branch and right surviving values. No separate repeated speedup claim belongs to this picture.

D06 · External task and tool



What runs

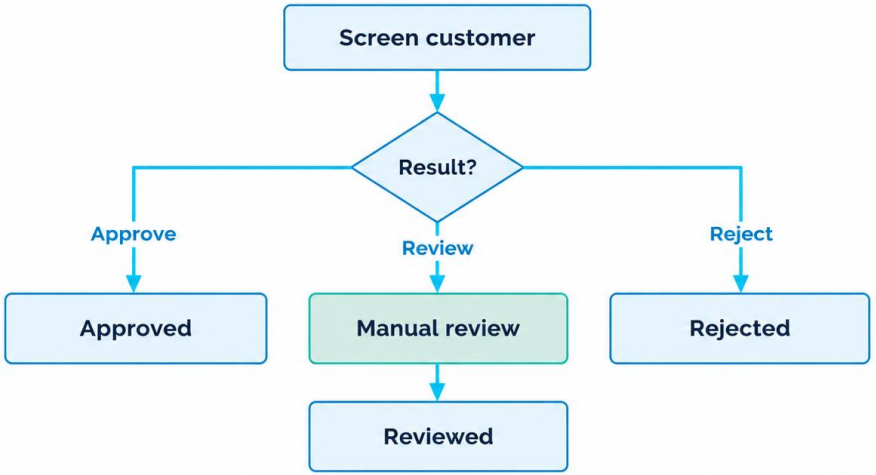
Fluxnova creates an external task. A worker fetches it, calls the deterministic echo tool and reports completion. The benchmark observes durable history.

This is the workload behind P01, the runtime/release comparisons and the soak. Their clocks include worker/tool work and observation.

D07 · KYC decision and review

BASIC WORKLOAD / KYC

Approve, review or reject?



Check every route, including the review wait.

Synthetic screening. The test simulates completing the review.

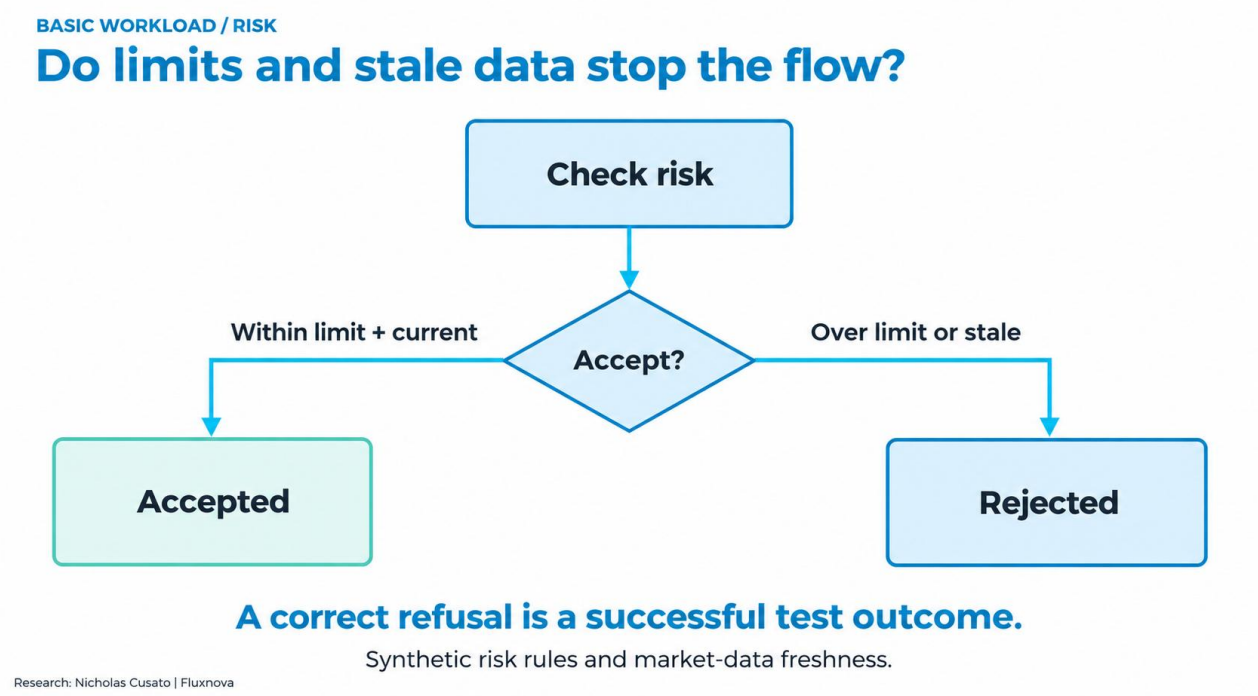
Research: Nicholas Cusato | Fluxnova

What runs

Synthetic screening approves, rejects or creates a review wait. The test simulates completion of that review.

A correct waiting or rejected state can be a passing test. Real screening quality and human working time remain outside this basic fixture.

D08 · Risk and freshness

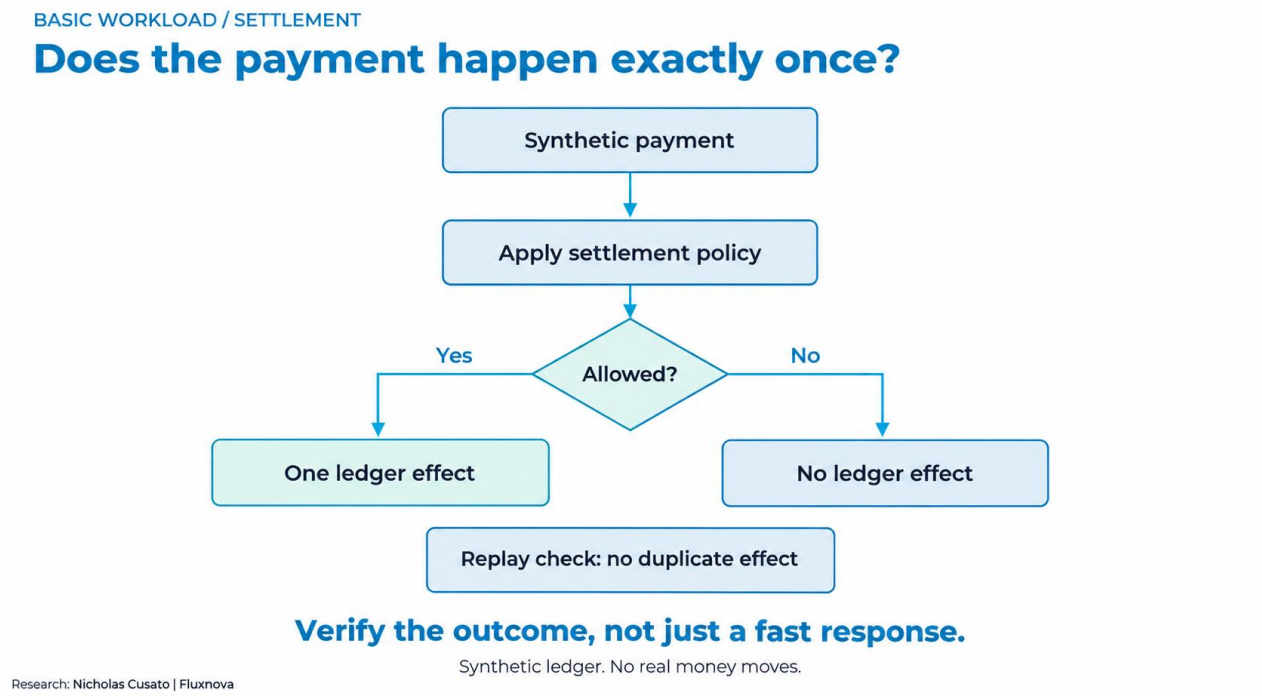


What runs

Check a synthetic risk limit and market-data freshness. Accept only when both declared conditions allow it.

An expected refusal is correct work, but it is not a settled payment. Report those populations separately.

D09 · Settlement and replay

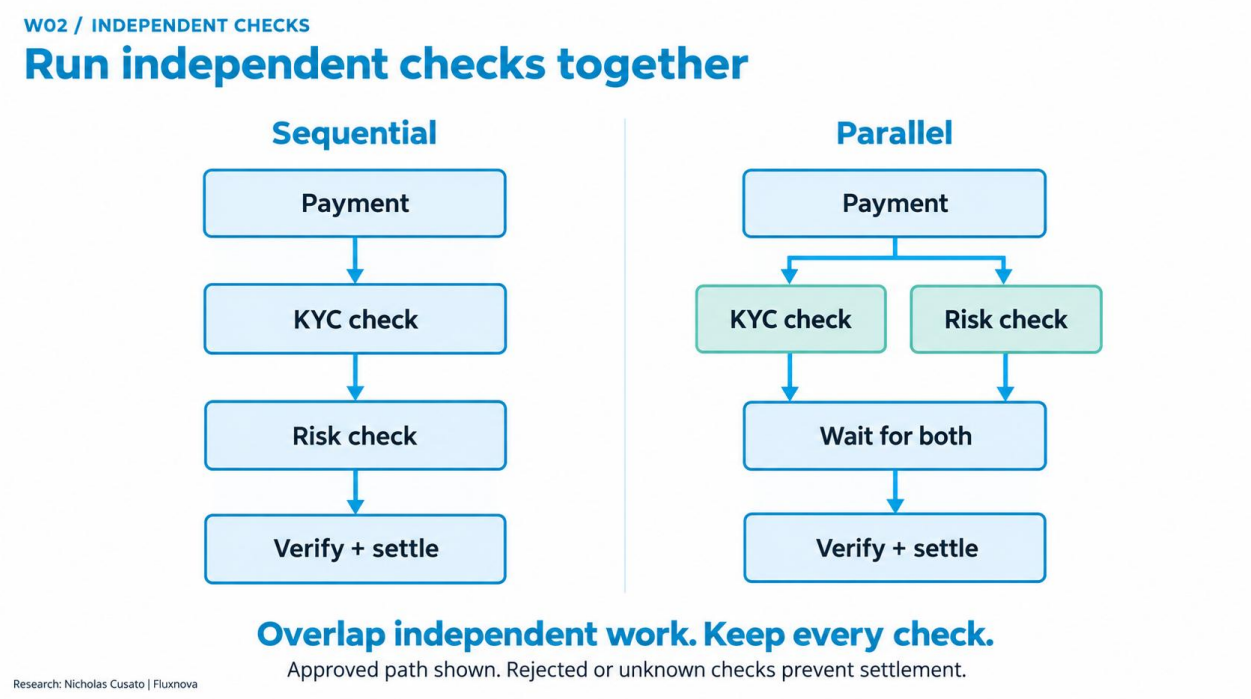


What runs

Check policy before the synthetic ledger effect. Repeating an operation must preserve its original result without paying again.

Successful HTTP responses are insufficient: the checker matches amount, beneficiary and effect count to the original intent. No real money moves.

Independent checks can overlap



32.1%

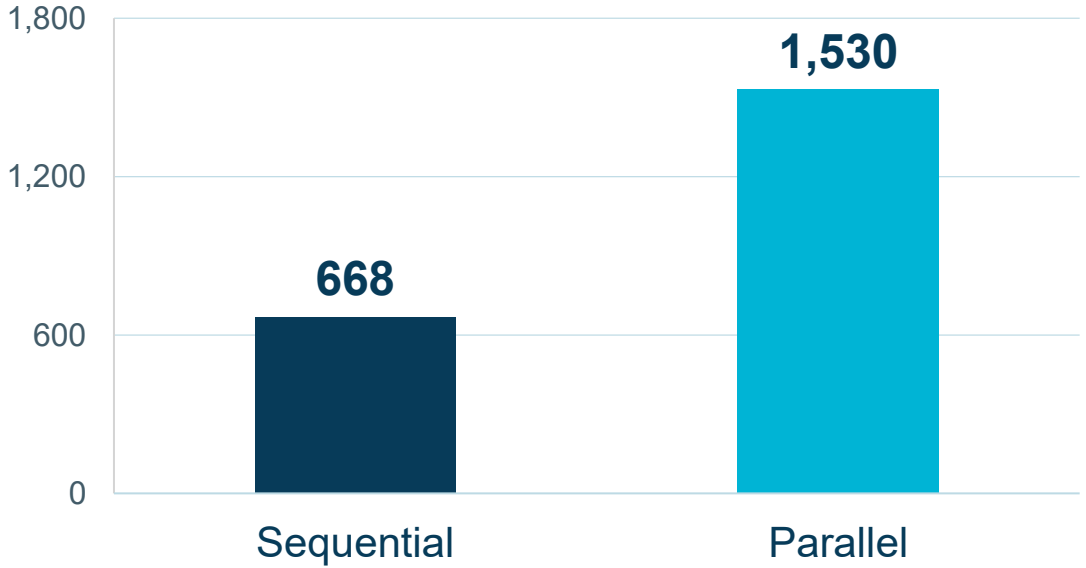
lower observed p95 at 1 payment/s

6,000 / 6,000 correct and on time

Both checks still finish before payment. Capacity was not measured.

Parallel FX quotes meet more deadlines

Correct settlements within two seconds



1,530 expected settlements per arm

3,960 / 3,960 correct outcomes

910 cached replays in the parallel arm
0 duplicate payments

Five pairs. Synthetic providers and delay profiles. Same three workers.

Runtime changes showed little difference

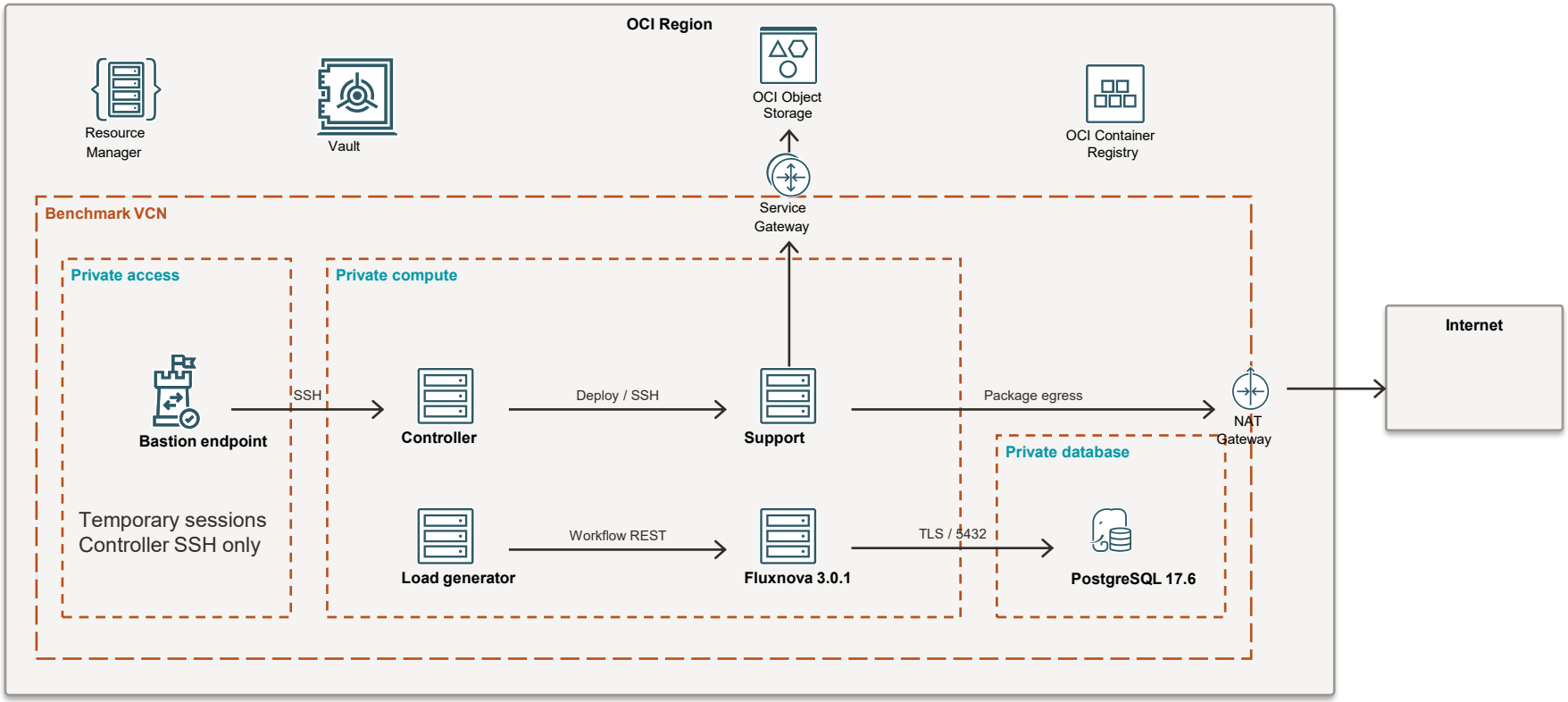
Both comparisons met the same fixed workload target

Separate comparison	Median paired p95 change	Range across five pairs
Java 25 vs Java 21	−0.862 ms	−1.369 to +0.864 ms
Fluxnova 3.1.1 vs 3.0.1	+0.126 ms	−0.552 to +1.055 ms

180,000 / 180,000 correct and on time in each study

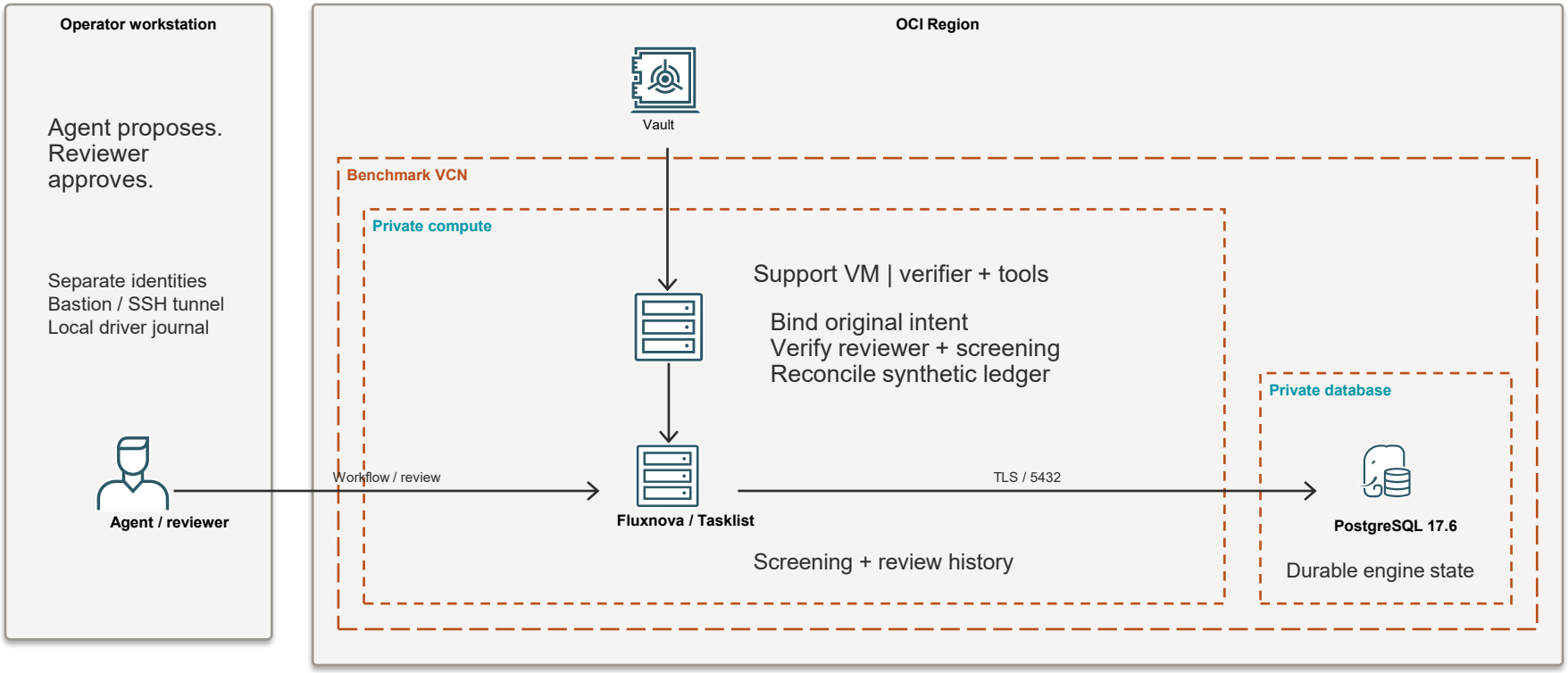
Small differences changed sign. No broad speedup was established.

OCI reference deployment



Original four-host reference. Other studies use their own declared setups.

Payment authority and review



Synthetic payment path. Model proposals require independent authority checks.