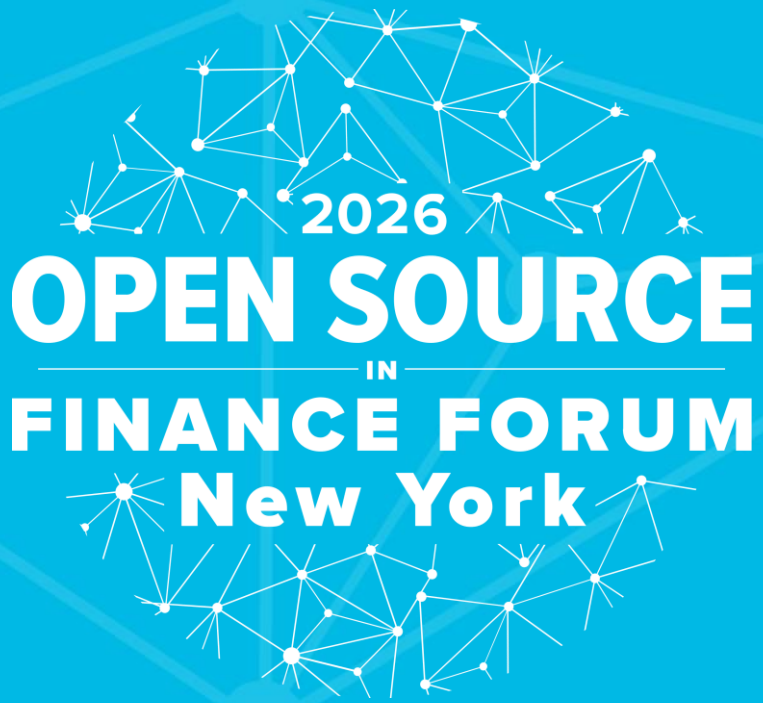


Benchmarking Fluxnova for Financial and Agentic Payment Workflows

Repeatable tests for performance, reliability and correct outcomes

Nicholas Cusato
Oracle Cloud Infrastructure

Introduction

Nicholas Cusato
Senior Cloud Engineer

Oracle Cloud Infrastructure

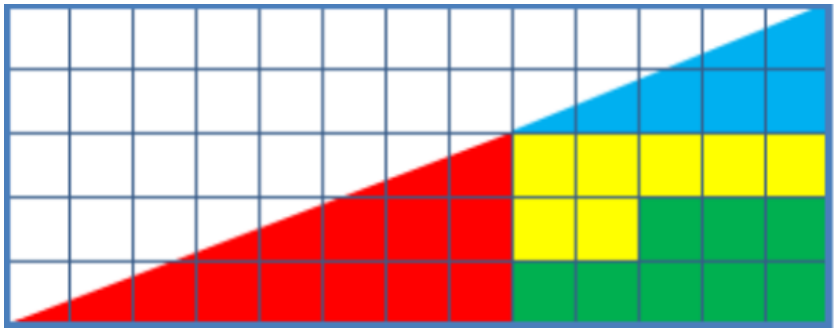
Financial Services and Insurance

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.**"

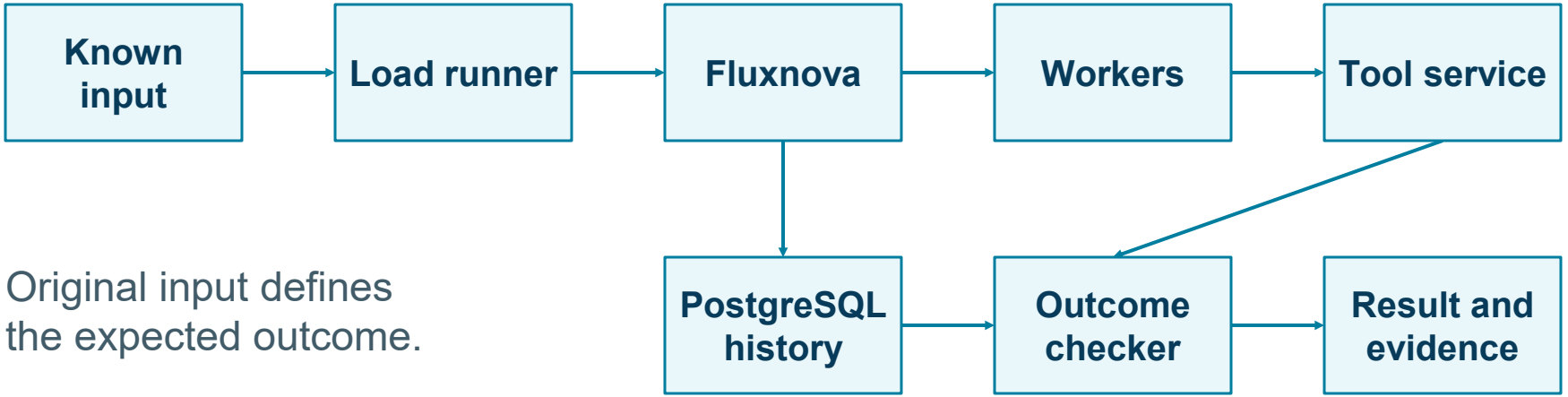
Make workflow changes with evidence

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

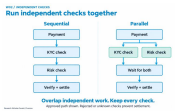
Confidence in workflow changes

6 areas	17 study records	
Workflow design	CORE	Basic flows
	W01	Worker pacing
	W02	Parallel checks
	W02 mixed	Mixed outcomes
	W03	Review routing
	PAY05	FX quotes
Runtime upgrades	JVM	Java 21 / 25
	RELEASE	Fluxnova releases
Load and endurance	P01	Reference baseline
	SOAK	Eight-hour run
	CAPACITY	Capacity (partial)
Events and recovery	FNV06	Events / batches
	FNV07	Retries / faults
	FNV07 timing	Timing (local)
	RECOVERY	State recovery
Agent controls	AGENT	Authority checks
Measurement	OBSERVER	Polling / timing

9 workflow diagrams



D01 Worker pacing



D02 Parallel checks



D03 Review routing



D04 Minimal flow



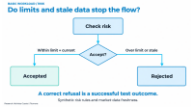
D05 Data / branching



D06 External work



D07 KYC / review



D08 Risk / freshness



D09 Settle / replay

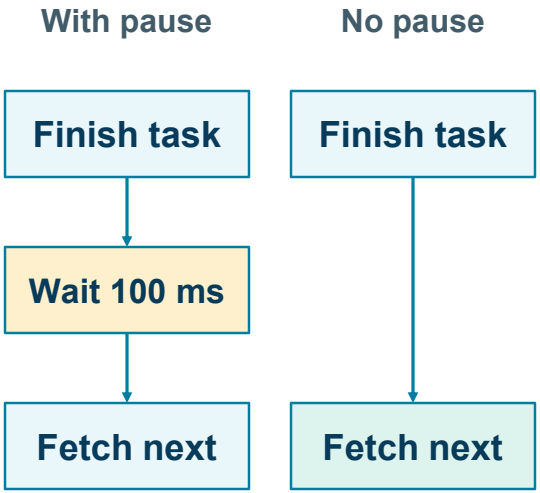
Explore the full catalogue

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

Completed, partial and diagnostic evidence

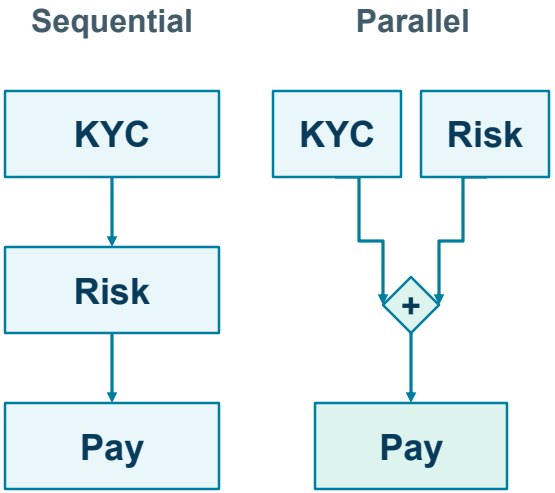
Three workflow comparisons

W01 Worker pacing



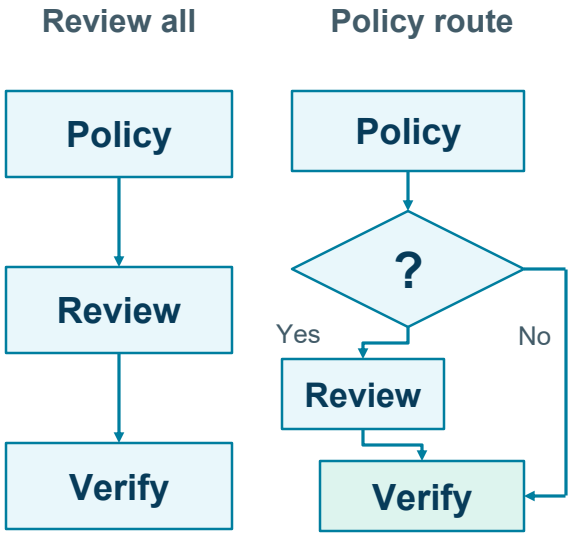
Fetch again after success.
Idle and retry backoff stays.

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.

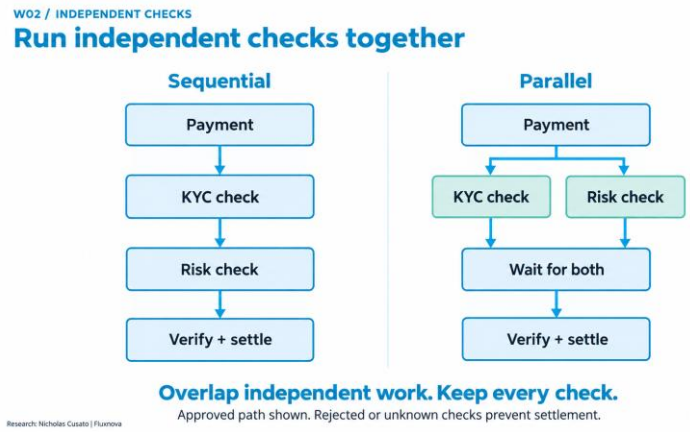
What the trials taught us

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

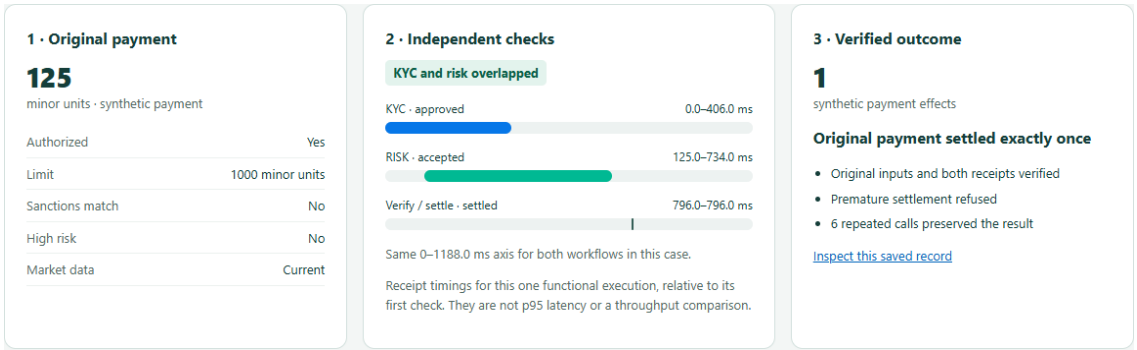
Validate other deployment adapters and compare declared alternatives.

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