AgentLand Observability

Air Traffic Control for AI Coding Agents

You are flying blind.

Running multiple AI coding agents simultaneously means tab-switching, scrolling through endless output, and hoping nothing breaks.

⚠ The 'Black Box' Problem:

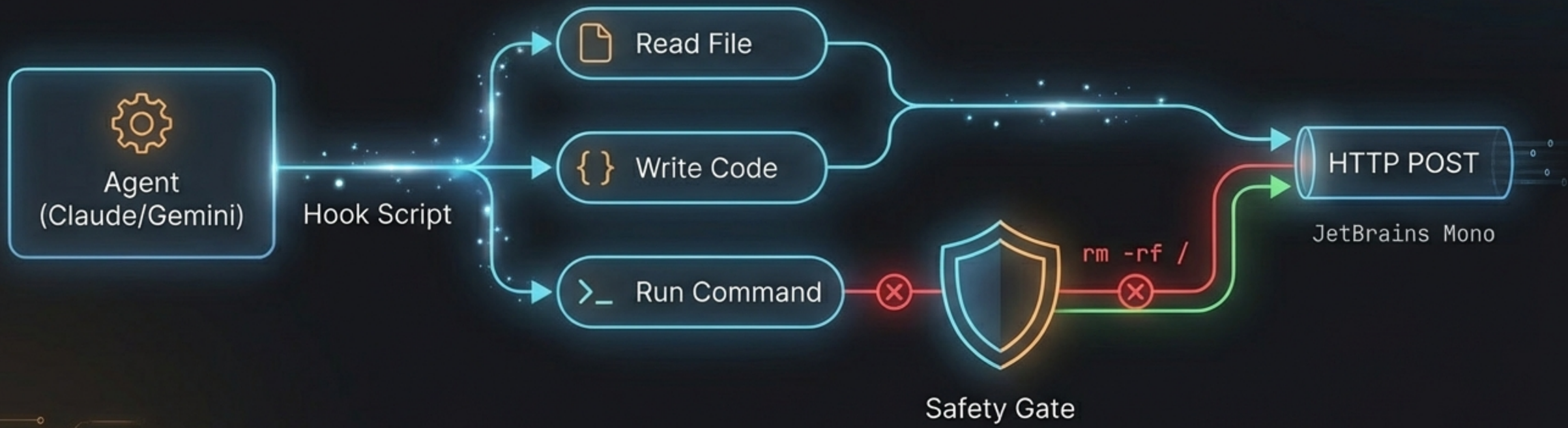
You don't know what decisions are being made until it's too late.

A Complete Observability Loop

The system captures every meaningful event, broadcasts it in real-time, visualizes the workflow, and judges the quality.

The Ears: Hooks & Safety

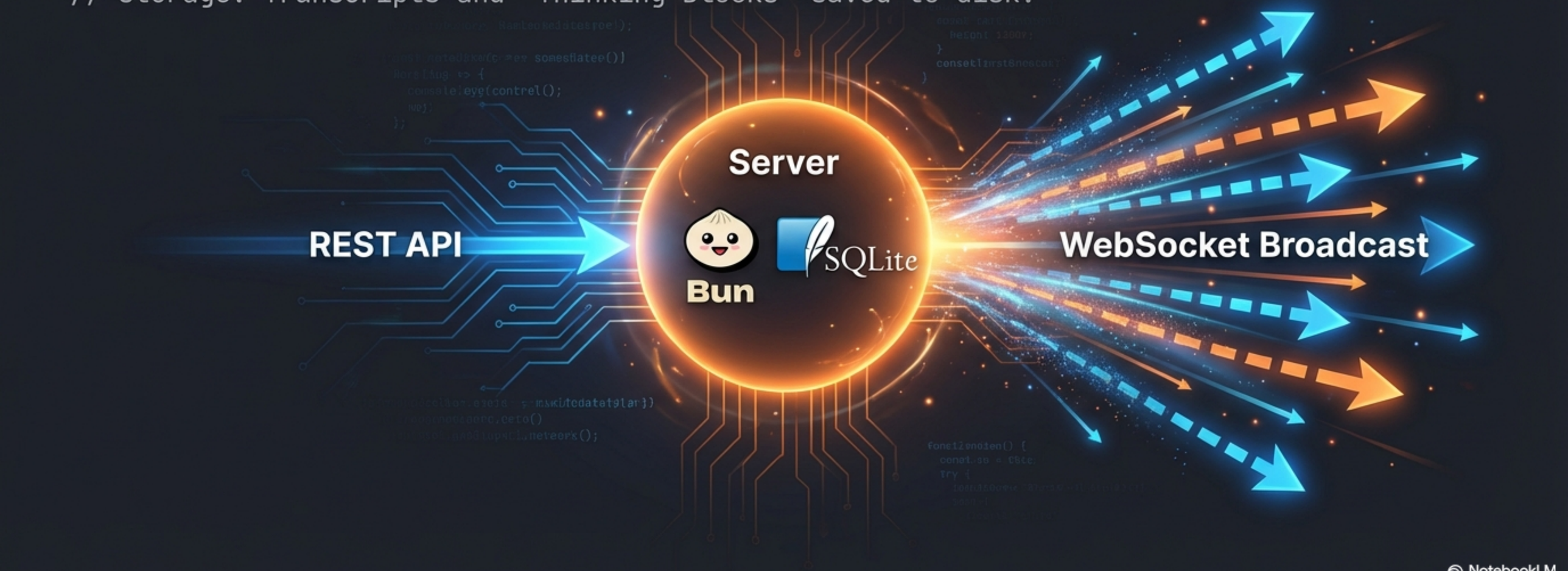
Python scripts plug directly into agent hooks. They capture events and fail silently—if the server is down, your agents keep working.



`JetBrains Mono`: Protocol: Lightweight HTTP POST delivery

The Brain: The Server

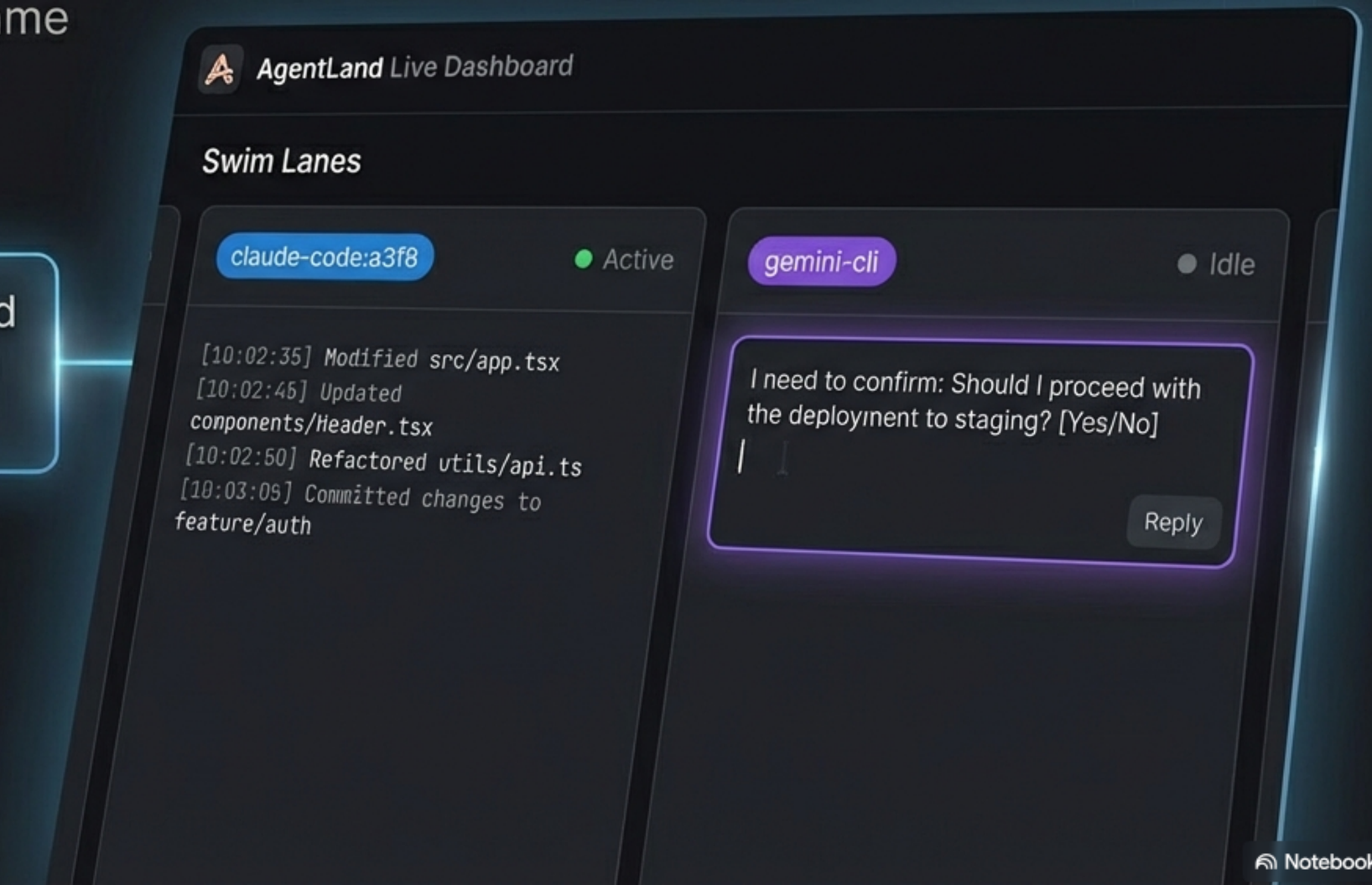
```
// In-Memory Speed: WAL mode for concurrent reads/writes.
// Real-Time: Immediate distribution to clients.
// Storage: Transcripts and 'Thinking Blocks' saved to disk.
```



The Eyes: Real-Time Visibility

Watch events stream instantly. Identify agents by Project Name or Source App.

Human-in-the-loop: Respond to agent questions directly in the feed.



Analytics & History

Move beyond the live feed. Analyze aggregate KPIs and audit full conversation histories stored permanently on disk.

Insights

Transcripts

User: What's the status of the build?

Agent: The build successfully completed at 10:45 AM. The build imis been successfullly completed at from agent-a.

▶ View Thinking Process

```
INTERNAL REASONING:
> Analyzing request type: Build Status Query.
> Checking active agents: agent-a (active), agent-b (idle).
> Tool Selection: Retrieve build logs from agent-a.
> Executing: get_logs('agent-a')
> Processing results: Status is SUCCESS.
```

Agent: The build successfully completed at 10:45 AM.

The Judge: Quality Control on Autopilot

Asynchronous evaluation runs in the background, scoring agent performance and flagging degradation automatically.

The Evaluators

Tool Success

JetBrains Mono

Pass/Fail rates

Regression Detection

JetBrains Mono

Z-score vs 7-day baseline

Transcript Quality

JetBrains Mono

LLM graded (1-5 scale)

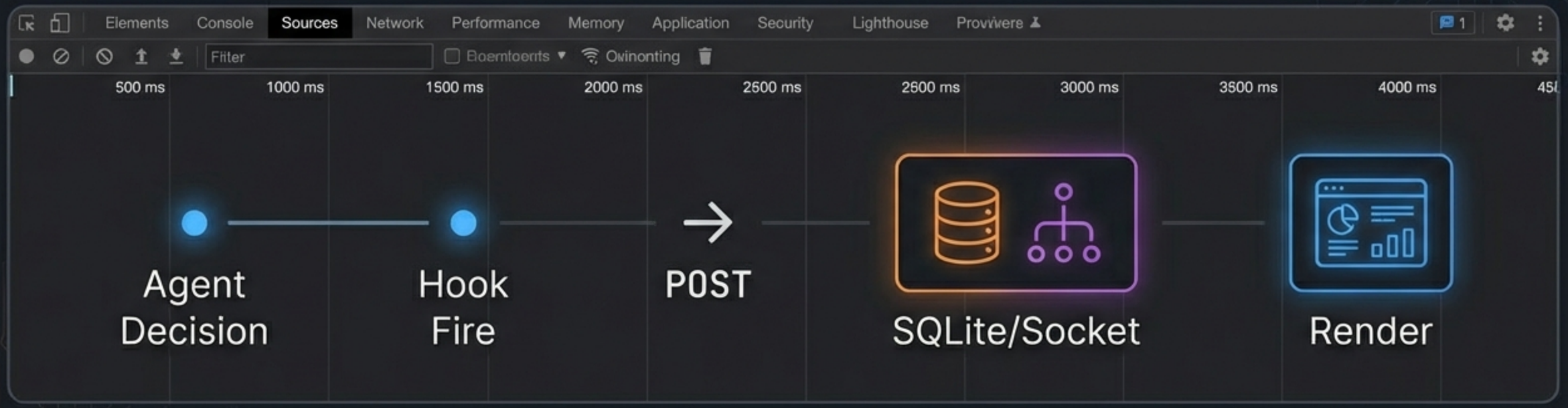
Reasoning Quality

JetBrains Mono

Depth & coherence check.

The Event Pipeline

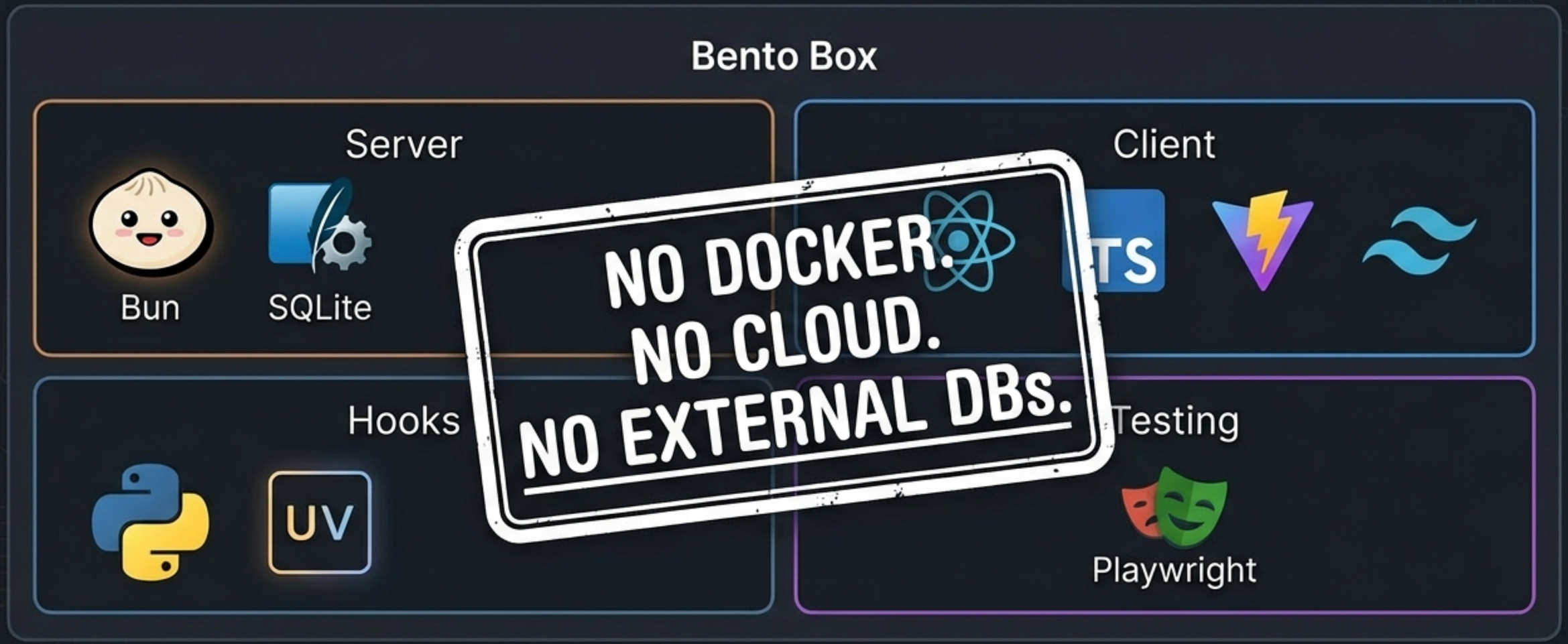
From agent decision to dashboard visualization in milliseconds.



< 50ms Total Latency

Deliberately Lightweight

Runs entirely locally on your machine. Zero external dependencies.



If you run agents on real codebases, you must watch what they do.

Move from hoping for the best to measuring tool failures, reasoning quality, and regression.

The Old Way (Opaque)

AgentLand (Measurable)

AgentLand Observability

CAPTURE. VISUALIZE. EVALUATE.

Control the chaos.