

To: ba5 (Lead Project Manager / Overseer)

From: ba6 (Verification & Monitoring Unit)

Date: 2025-05-30

Subject: Detailed Findings Report - Tasks 0.5.1 & 0.5.2: Performance Benchmarking Implementation and Initial Baseline Data (Addendum on Test Limitations)

(Previous sections of the report remain as is)

Addendum: Limitations and Observations from Initial Short Test Runs

ba5, while the instrumentation and data dumping mechanisms are now functional, it's important to highlight specific characteristics and limitations observed during the brief (e.g., 10-second) test runs conducted to verify the setup:

1. Zero or Near-Zero Timings for Some Metrics:

- In some early test runs (before "actionable signals" in the video feed), metrics like `eventbus.handler_time.OCRParsedEvent._handle_ocr_parsed` and `ocr.conditioning_time_ms` reported all statistical values (mean, median, max, min) as 0.0000 ms, despite having a non-zero count.
- **Interpretation:** This strongly suggests that for the specific (likely very simple or minimal) OCR data being processed in those initial short tests, the actual execution time of the instrumented code sections was below the effective resolution of `time.time()` on the test system, or the operations were genuinely near-instantaneous.
- **Impact:** While this doesn't indicate a failure of the benchmarker, it means that for extremely fast operations, the current `time.time()` based measurement might not provide granular sub-millisecond insight. For baselining these ultra-fast paths, longer runs averaging more samples or potentially using `time.perf_counter()` (if deemed necessary later) might be considered for higher precision. However, for operations that take even a few milliseconds, the current setup proved adequate in subsequent tests.

2. High Latency & Variance

in `signal_gen.ocr_cleaned_to_order_request_latency_ms`:

- The "actionable signals" test run (Count: 6) showed a mean of ~474ms, but a median of only ~9ms, with p95/max values exceeding 1.3-1.5 seconds.

- **Interpretation:** This indicates a long-tail distribution. The majority of signal generations are fast, but a few instances are significantly slower.
- **Contextual Cause (After-Hours):** Given the tests were run during after-hours, this slowness in the tail is almost certainly attributable to OCRScalpingSignalOrchestratorService relying on API fallbacks via IPriceProvider to fetch price data (e.g., for LMT order calculations or strategy context). API calls are inherently much slower than accessing streamed or cached prices.
- **Value of this Data:** This provides a useful "worst-case" or "after-hours/API-dependent" baseline for this specific latency. It highlights a known bottleneck when live, low-latency price streams aren't available.

3. Inactivity of Price-Related Paths (pmd.* metrics):

- Metrics
for PriceMeltdownDetector (pmd.process_trade_time_ms, pmd.process_quote_time_ms) consistently showed "No data collected."
- **Interpretation:** This is expected. The tests focused on OCR video replay. Without a simulated or live market data feed generating MarketDataTickEvent and MarketDataQuoteEvent during these short after-hours tests, the PMD would not be invoked.
- **Implication:** Baselines for these components and the market data event bus path are still pending and require tests under RTH conditions or with a market data simulator.

4. Test Duration and Sample Size:

- The test runs providing data were very short (e.g., the last one producing 11 OCRParsedEvents and 6 OrderRequestEvents).
- **Interpretation:** While sufficient to verify functionality and get initial rough numbers, these short durations provide a limited sample size. For more statistically robust baselines, longer runs (e.g., 15-30 minutes or more, processing a larger volume of varied data) will be necessary. The current numbers should be treated as preliminary indicators.

Summary of Limitations for ba5:

The current baseline numbers are preliminary and heavily influenced by:

- The after-hours testing environment (impacting price data availability and latency).
- The short duration of the verification test runs (potentially affecting the statistical significance for some metrics and the granularity of measurements for very fast operations).
- The nature of the input (video replay for OCR, no active market data simulation for PMD).

These limitations are understood, and the plan is to address them by conducting more comprehensive baseline runs during active market hours and/or with more extensive simulated data feeds. The current setup, however, has proven capable of capturing the required data points.

(End of Addendum for ba5)

22.2s

edit

more_vert

edit

more_vert

edit

more_vert

edit

more_vert