

MEMORANDUM: DOM-ASM Single-Pass Architecture Migration

TO: Aegis Development Team
FROM: Engineering Architecture Team
DATE: January 2025
RE: HTML/CSS Pipeline Migration to Single-Pass Architecture

EXECUTIVE SUMMARY

This memorandum outlines the technical implementation roadmap for converting DOM-ASM HTML and CSS components to unified single-pass architecture (`TOKENIZER → PARSER → AST`). The migration implements Ship of Theseus behavioral equivalence principles while achieving measurable performance optimization through state minimization.

CRITICAL CONSTRAINT: JavaScript pipeline development remains frozen until HTML and CSS achieve production stability.

MILESTONE FRAMEWORK

Phase 1A: HTML Pipeline Unification

Duration: 4-5 sprints (8-10 weeks)
Resource Allocation: 2 senior engineers + 1 QA engineer
Cost Equivalence Target: 350-420 development hours

Milestone 1.1: Core Pipeline Architecture (Week 1-2)

Deliverables:

- ☐ Consolidate `src/html/tokenization/` + `src/html/parsing/` → `src/html/pipeline/`
- ☐ Implement unified `HTMLTokenizer.ts` with linear token stream generation
- ☐ Develop `HTMLParser.ts` eliminating circular parsing dependencies
- ☐ Create `HTMLASTBuilder.ts` integrating existing state minimization algorithms

Success Criteria:

- ☐ TypeScript compilation across unified interfaces
- ☐ Zero recursive function calls in token → AST pipeline
- ☐ Memory allocation reduction of minimum 15% compared to multi-pass baseline
- ☐ Linear O(n) complexity validation through performance profiling

Cost Validation:

- Estimated: 140-170 hours

- Acceptance Threshold: <180 hours actual development time

Milestone 1.2: AST Integration and Optimization (Week 3)

Deliverables:

- ☐ Integrate existing `HTMLeOptimizer.js` patterns into single-pass flow
- ☐ Implement behavioral equivalence validation framework
- ☐ Develop state machine minimization within AST construction phase
- ☐ Create comprehensive test suite for Ship of Theseus compliance

Success Criteria:

- ☐ AST output functionally identical to existing multi-pass implementation
- ☐ State minimization achieving equivalent optimization to current system
- ☐ Automated behavioral regression testing infrastructure operational
- ☐ Performance improvement of minimum 20% in parsing throughput

Cost Validation:

- Estimated: 40-50 hours
- Acceptance Threshold: <60 hours actual development time

Milestone 1.3: Interface Standardization (Week 4)

Deliverables:

- ☐ Implement unified `Core.compile(input: string, format: 'html')` interface
- ☐ Create comprehensive type definitions in `src/html/types/`
- ☐ Develop validation utilities for behavioral equivalence checking
- ☐ Document migration path for existing adopters

Success Criteria:

- ☐ Complete TypeScript type safety across all HTML components
- ☐ CLI integration functional with unified interface
- ☐ Backward compatibility maintained for existing API consumers
- ☐ Documentation complete with code examples and migration guides

Cost Validation:

- Estimated: 30-40 hours
- Acceptance Threshold: <50 hours actual development time

Phase 1B: CSS Pipeline Implementation

Duration: 6-7 sprints (12-14 weeks)

Resource Allocation: 3 senior engineers + 1 CSS specialist + 1 QA engineer

Cost Equivalence Target: 450-560 development hours

Milestone 2.1: CSS Tokenization Architecture (Week 5-6)

Deliverables:

- ☐ Design CSS tokenizer from scratch with single-pass constraints
- ☐ Implement selector tokenization preserving specificity calculations
- ☐ Develop media query and pseudo-class handling optimization
- ☐ Create CSS token type definitions and validation framework

Success Criteria:

- ☐ Complete CSS 3.0 specification compliance in tokenization
- ☐ Selector specificity preservation with mathematical validation
- ☐ Cross-browser compatibility testing framework operational
- ☐ Performance target: $O(n)$ complexity for stylesheet tokenization

Cost Validation:

- Estimated: 100-120 hours
- Acceptance Threshold: <140 hours actual development time

Milestone 2.2: CSS Parser and Cascade Logic (Week 7-9)

Deliverables:

- ☐ Implement cascade resolution algorithm within single-pass constraints
- ☐ Develop rule precedence calculation maintaining CSS specification compliance
- ☐ Create stylesheet AST construction with property optimization
- ☐ Build cross-browser rendering validation system

Success Criteria:

- ☐ Cascade behavior identical to browser native implementations
- ☐ Property value normalization without semantic loss
- ☐ Computed style calculations achieving specification compliance
- ☐ Performance improvement of minimum 25% over multi-pass parsing

Cost Validation:

- Estimated: 180-200 hours
- Acceptance Threshold: <220 hours actual development time

Milestone 2.3: CSS-HTML Integration (Week 10-11)

Deliverables:

- ☐ Integrate CSS AST with HTML AST for unified DOM representation
- ☐ Implement style application algorithms within single-pass flow
- ☐ Create comprehensive cross-component validation testing
- ☐ Develop performance benchmarking against existing implementations

Success Criteria:

- ☐ Unified DOM-CSS AST achieving complete behavioral equivalence
- ☐ Style application performance matching or exceeding browser baselines
- ☐ Memory efficiency improvement of minimum 30% through unified representation
- ☐ Complete regression testing suite operational

Cost Validation:

- Estimated: 80-100 hours
 - Acceptance Threshold: <120 hours actual development time
-

COST EQUIVALENCE VALIDATION FRAMEWORK

Performance Benchmarking Requirements

Baseline Metrics (Pre-Migration):

- ☐ Current HTML parsing throughput measurement
- ☐ Existing CSS cascade resolution performance profiling
- ☐ Memory allocation patterns documentation
- ☐ Multi-pass architecture computational overhead analysis

Target Performance Improvements:

- ☐ HTML pipeline: Minimum 20% throughput improvement
- ☐ CSS pipeline: Minimum 25% parsing performance enhancement
- ☐ Combined memory usage: Maximum 30% reduction
- ☐ State machine optimization: Equivalent or superior minimization results

Behavioral Equivalence Testing

Ship of Theseus Compliance Validation:

- ☐ Input/output functional testing across 10,000+ HTML document samples
 - ☐ CSS cascade behavior verification against W3C test suites
 - ☐ Cross-browser rendering consistency validation
 - ☐ State transition equivalence mathematical proof documentation
-

RISK MITIGATION PROTOCOLS

High-Risk Technical Factors

HTML Pipeline Risks:

- ☐ Behavioral regression during tokenizer unification
- ☐ Performance degradation in AST optimization integration
- ☐ Interface compatibility breaking existing adopter implementations

CSS Pipeline Risks:

- ☐ Cascade behavior modification during single-pass conversion
- ☐ Selector specificity calculation errors
- ☐ Cross-browser compatibility regressions

Mitigation Strategies

- ☐ Comprehensive automated regression testing at each milestone
- ☐ Performance monitoring with automatic rollback triggers
- ☐ Staged deployment with backward compatibility maintenance
- ☐ Weekly stakeholder progress reviews with technical validation

RESOURCE ALLOCATION SUMMARY

Total Project Investment

HTML Pipeline: 350-420 hours (Risk-adjusted: 420-500 hours)

CSS Pipeline: 450-560 hours (Risk-adjusted: 560-670 hours)

Combined Total: 800-980 hours (Risk-adjusted: 980-1170 hours)

Timeline Summary

Phase 1A (HTML): Weeks 1-4 (1 month)

Phase 1B (CSS): Weeks 5-11 (1.75 months)

Total Duration: 11 weeks (2.75 months)

Success Metrics

- ☐ Zero behavioral regressions in DOM manipulation functionality
- ☐ Minimum 20% overall performance improvement
- ☐ Complete TypeScript type safety across unified architecture
- ☐ Successful CLI integration with `Core.compile()` interface

APPROVAL REQUIREMENTS

Technical Approval: Senior Engineering Manager

Resource Approval: Development Team Lead

Timeline Approval: Project Management Office

Next Action: Stakeholder review and resource allocation authorization for Phase 1A initiation.

Document Control: Version 1.0 | Classification: Internal Technical Documentation