Table of Contents

# 🚀 Implementation Plan - 4 Weeks (2 Sprints)

## Overview

**Timeline:** 4 weeks (2 sprints) **Team:** 1 SM, 1 PO, 2 Data Engineers, 1 Tester **Approach:** Agile, parallel work where possible

## Team Composition & Roles

| **Role** | **Count** | **Responsibilities** |
| --- | --- | --- |
| **Scrum Master** | 1 | Facilitation, impediment removal, sprint ceremonies |
| **Product Owner** | 1 | Requirements validation, acceptance testing, stakeholder communication |
| **Data Engineer** | 2 | Code adaptation, deployment, performance optimization |
| **Tester** | 1 | Test execution, validation, quality assurance |

## Sprint 1: Setup & Initial Implementation (Week 1-2)

### Week 1: Understanding & Setup

#### Day 1: Kickoff & Onboarding

**All Team (4 hours)** - Morning: Sprint Planning & Kickoff (2h) - Review repository structure (<INDEX.md>) - Assign responsibilities - Setup environments - Afternoon: Individual deep-dives (2h)

**Parallel Work:**

| **Role** | **Task** | **Time** | **Deliverable** |
| --- | --- | --- | --- |
| **PO** | Read Executive Summary + Scenarios Guide | 2h | Requirements validated |
| **Engineer 1** | Review main code ([databricks\_video\_aggregation.py](03_DEVELOPMENT/databricks_video_aggregation.py)) | 2h | Understanding of logic |
| **Engineer 2** | Setup Databricks workspace & permissions | 2h | Environment ready |
| **Tester** | Read test specification ([TEST\_SCENARIOS\_COMPLETE.md](04_TESTING/TEST_SCENARIOS_COMPLETE.md)) | 2h | Test plan ready |
| **SM** | Setup Jira/ADO tickets & sprint board | 2h | Sprint backlog |

#### Day 2-3: Adaptation & Configuration

**Focus:** Adapt code to company infrastructure

**Engineer 1 Tasks (2 days):** - [ ] Adapt table names to company schema (2h) - [ ] Configure input/output table paths (1h) - [ ] Update video metadata source (1h) - [ ] First test run with sample data (2h) - [ ] Document configuration changes (1h) - [ ] Code review with Engineer 2 (1h)

**Engineer 2 Tasks (2 days):** - [ ] Create video metadata table (2h) - [ ] Setup Delta Lake tables (2h) - [ ] Configure cluster settings (1h) - [ ] Setup job scheduling (2h) - [ ] Create monitoring dashboard skeleton (1h)

**Tester Tasks (2 days):** - [ ] Setup test environment in Databricks (2h) - [ ] Generate test data using notebook (2h) - [ ] Create test data table (1h) - [ ] Prepare validation queries (2h) - [ ] Setup test result tracking (1h)

**PO Tasks (2 days):** - [ ] Validate business requirements against scenarios (3h) - [ ] Identify company-specific edge cases (2h) - [ ] Draft acceptance criteria (2h) - [ ] Stakeholder alignment meeting (1h)

#### Day 4-5: First Integration & Smoke Tests

**Focus:** Get end-to-end working

**All Engineers (2 days):** - [ ] Run aggregation on test data (3h) - [ ] Debug issues (4h) - [ ] Validate output schema (2h) - [ ] Performance baseline measurement (1h) - [ ] Documentation of changes (2h)

**Tester (2 days):** - [ ] Smoke test: TC-001 (Perfect Viewing) (30 min) - [ ] Smoke test: TC-002 (Pause & Resume) (30 min) - [ ] Smoke test: TC-003 (Browser Close) (30 min) - [ ] Validate basic metrics correctness (2h) - [ ] Log issues in Jira/ADO (1h) - [ ] Create initial test report (1h)

**Daily Standup (15 min × 5 days = 1.25h)**

**Sprint Review Week 1 (1h - Friday afternoon)**

### Week 2: Core Scenarios & Refinement

#### Day 6-7: Core Testing & Bug Fixes

**Focus:** Validate all P0 scenarios

**Tester (2 days):** - [ ] **P0 Core Tests (6 scenarios):** (4h) - TC-001: Perfect Viewing - TC-002: Pause & Resume - TC-003: Browser Close - TC-013: Null/Missing Values - TC-004: Skip Forward - TC-005: Skip Backward - [ ] Run validation queries (1h) - [ ] Document bugs (2h) - [ ] Retest fixes (2h)

**Engineers (2 days - parallel to testing):** - [ ] Fix P0 bugs as they come in (6h) - [ ] Code optimization (4h) - [ ] Add missing data quality checks (2h)

**PO (2 days):** - [ ] Review first results (2h) - [ ] Validate business logic (3h) - [ ] Prepare for demo (1h)

#### Day 8-10: Advanced Testing & Dashboard

**Focus:** Complete testing, build dashboard

**Tester (3 days):** - [ ] **P1 Tests (9 scenarios):** (4h) - Gaming detection, multi-session, etc. - [ ] **P2 Sample Tests (5 of 10):** (2h) - Edge cases, boundary conditions - [ ] Final validation queries (1h) - [ ] Create comprehensive test report (2h) - [ ] Sign-off meeting preparation (1h)

**Engineers (3 days):** - [ ] Fix remaining bugs (4h) - [ ] Connect to BI tool (Power BI/Tableau) (6h) - [ ] Create initial dashboards (6h) - [ ] Performance tuning (4h)

**PO (3 days):** - [ ] Dashboard review & feedback (4h) - [ ] User acceptance testing (4h) - [ ] Prepare stakeholder demo (2h)

**Sprint 1 Review & Retrospective (2h - Friday afternoon)**

## Sprint 2: Production Deployment & Optimization (Week 3-4)

### Week 3: Production Preparation

#### Day 11-12: Production Deployment

**Focus:** Deploy to production environment

**Engineers (2 days):** - [ ] Create production tables (2h) - [ ] Configure production job (3h) - [ ] Setup monitoring & alerts (3h) - [ ] Run first production aggregation (2h) - [ ] Validate production data (2h) - [ ] Document production setup (2h)

**Tester (2 days):** - [ ] Validate production data (4h) - [ ] Spot-check random samples (2h) - [ ] Compare dev vs prod results (2h) - [ ] Final quality report (2h)

**PO (2 days):** - [ ] Stakeholder communication (2h) - [ ] Prepare Go-Live checklist (2h) - [ ] Risk assessment (2h)

#### Day 13-15: Dashboard Finalization & Training

**Focus:** Finish dashboards, train users

**Engineers (3 days):** - [ ] Finalize all dashboard views (6h) - [ ] Add drill-down capabilities (4h) - [ ] Performance optimization (4h) - [ ] Create dashboard documentation (2h)

**Tester (3 days):** - [ ] UAT with real users (4h) - [ ] Dashboard testing (3h) - [ ] Performance testing (3h) - [ ] Final sign-off (1h)

**PO (3 days):** - [ ] User training sessions (6h) - [ ] Documentation review (2h) - [ ] Stakeholder demo (2h) - [ ] Collect feedback (2h)

### Week 4: Monitoring, Optimization & Handover

#### Day 16-18: Monitoring & Fine-tuning

**Focus:** Ensure stability, optimize based on real data

**Engineers (3 days):** - [ ] Monitor production runs (6h) - [ ] Optimize based on data volume (4h) - [ ] Fix any production issues (4h) - [ ] Create runbook (2h)

**Tester (3 days):** - [ ] Monitor data quality (4h) - [ ] Validate ongoing aggregations (3h) - [ ] Update test suite if needed (2h)

**PO (3 days):** - [ ] Collect user feedback (4h) - [ ] Prioritize enhancements (2h) - [ ] Plan Phase 2 features (4h)

#### Day 19-20: Handover & Retrospective

**Focus:** Knowledge transfer, documentation, retrospective

**All Team (2 days):** - [ ] Knowledge transfer sessions (4h) - [ ] Documentation finalization (3h) - [ ] Sprint 2 Review (2h) - [ ] Sprint 2 Retrospective (1.5h) - [ ] Celebration! (30 min)

**Final Deliverables:** - [ ] Production-ready aggregation pipeline - [ ] Interactive dashboards - [ ] Complete documentation - [ ] Runbook for operations - [ ] Test report & sign-off - [ ] Lessons learned document

## Effort Summary

### Total Effort by Role (4 weeks)

| **Role** | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Total** |
| --- | --- | --- | --- | --- | --- |
| **Scrum Master** | 20h | 20h | 20h | 20h | **80h** (2 FTE weeks) |
| **Product Owner** | 24h | 28h | 26h | 20h | **98h** (2.5 FTE weeks) |
| **Engineer 1** | 32h | 36h | 32h | 28h | **128h** (3.2 FTE weeks) |
| **Engineer 2** | 32h | 36h | 32h | 28h | **128h** (3.2 FTE weeks) |
| **Tester** | 24h | 28h | 26h | 18h | **96h** (2.4 FTE weeks) |
| **Total** | **132h** | **148h** | **136h** | **114h** | **530h** |

### Effort by Phase

| **Phase** | **Duration** | **Effort** | **Key Activities** |
| --- | --- | --- | --- |
| **Sprint 1 - Week 1** | 5 days | 132h | Setup, adaptation, smoke tests |
| **Sprint 1 - Week 2** | 5 days | 148h | Core testing, bug fixes, dashboard start |
| **Sprint 2 - Week 3** | 5 days | 136h | Production deployment, finalization |
| **Sprint 2 - Week 4** | 5 days | 114h | Monitoring, optimization, handover |
| **Total** | **20 days** | **530h** | **Complete implementation** |

## Critical Path

Week 1: Setup & Adaptation  
├─ Day 1: Kickoff  
├─ Day 2-3: Code adaptation (CRITICAL)  
├─ Day 4-5: First integration (CRITICAL)  
└─ Milestone: Working aggregation on test data  
  
Week 2: Testing & Refinement  
├─ Day 6-7: P0 testing (CRITICAL)  
├─ Day 8-10: Advanced testing + Dashboard  
└─ Milestone: All P0/P1 tests pass  
  
Week 3: Production Deployment  
├─ Day 11-12: Prod deployment (CRITICAL)  
├─ Day 13-15: Dashboard finalization  
└─ Milestone: Production ready  
  
Week 4: Stability & Handover  
├─ Day 16-18: Monitoring & tuning  
├─ Day 19-20: Handover  
└─ Milestone: Go-Live ✅

## Risk Management

### High Risks

| **Risk** | **Probability** | **Impact** | **Mitigation** |
| --- | --- | --- | --- |
| Data structure mismatch | Medium | High | Early validation in Week 1 Day 1-2 |
| Performance issues with real data | Low | High | Baseline testing in Week 1, optimization in Week 3-4 |
| Missing requirements | Low | Medium | PO validates requirements in Week 1 |
| Key team member unavailable | Low | High | Knowledge sharing, pair programming |

### Medium Risks

| **Risk** | **Probability** | **Impact** | **Mitigation** |
| --- | --- | --- | --- |
| BI tool integration complex | Medium | Medium | Allocate extra time in Week 2-3 |
| Edge cases not covered | Low | Medium | Comprehensive test suite (25 scenarios) |
| Stakeholder feedback late | Medium | Low | Early demos, continuous communication |

## Success Criteria

### Sprint 1 (Week 1-2)

* Aggregation runs successfully on test data
* All P0 test scenarios pass (6 scenarios)
* At least 80% of P1 scenarios pass (7 of 9)
* Basic dashboard shows correct data
* No critical bugs
* Code reviewed and approved

### Sprint 2 (Week 3-4)

* Production deployment successful
* All P0 and P1 scenarios pass in production (15 total)
* Dashboards complete and validated
* Performance meets SLA (queries < 15s)
* Stakeholder sign-off received
* Documentation complete
* Knowledge transfer completed

## Timeline Assumptions

### What’s Included in 4 Weeks

✅ Code adaptation to company infrastructure ✅ Full testing (P0, P1, sample P2) ✅ Dashboard creation (basic + advanced views) ✅ Production deployment ✅ Monitoring setup ✅ User training ✅ Documentation ✅ Knowledge transfer

### What’s NOT Included

❌ Major architectural changes ❌ Custom features beyond core requirements ❌ Integration with other systems (beyond BI tools) ❌ Real-time streaming (batch only) ❌ Advanced ML/predictive analytics

### Assumptions

1. **Code baseline is solid** - The provided solution works as-is
2. **Data structure similar** - Raw events have required fields (userId, videoId, timestamp, eventName, currentTime)
3. **Team availability** - All team members at least 80% dedicated
4. **Infrastructure ready** - Databricks environment accessible
5. **Stakeholder availability** - Available for reviews and sign-offs
6. **No major blockers** - Permissions, approvals move quickly

## Optimization Opportunities

If timeline needs to compress to **3 weeks:**

**Cut these low-priority items:** - ⬇️ P2 edge case testing (save 1 day) - ⬇️ Advanced dashboard features (save 1 day) - ⬇️ Week 4 monitoring period (reduce to 3 days, save 2 days) - ⬇️ Formal training sessions (async documentation instead, save 1 day)

**Total savings: 5 days → 3 weeks possible** ⚠️ Higher risk

## Post-Implementation (Optional Phase 2)

After 4-week initial implementation, consider:

**Future Enhancements (Backlog):** - Real-time streaming aggregation (Structured Streaming) - Predictive analytics (completion probability) - A/B testing framework - Advanced segmentation (cohort analysis) - Mobile vs Desktop tracking - Video quality tracking - Automated anomaly detection

**Timeline for Phase 2:** 2-4 weeks (depending on scope)

## Communication Plan

### Daily

* 15-min standup (blockers, progress, plan)
* Slack updates on critical issues

### Weekly

* Sprint review (Friday, 1h)
* Sprint planning (Monday, 1h for Sprint 2)
* Demo to stakeholders (Friday, 30 min)

### Bi-weekly

* Sprint retrospective (Friday, 1.5h)

### Milestone-based

* Go/No-Go decision after Week 2
* Production go-live approval after Week 3
* Final sign-off after Week 4

## Next Steps

1. **Week 0 (Preparation):**
   * Schedule kickoff meeting
   * Ensure all team members have repository access
   * Setup Databricks workspace
   * Create sprint board (Jira/ADO)
   * Identify stakeholders
2. **Sprint 1 Day 1:**
   * Team reads <INDEX.md>
   * Sprint planning meeting
   * Environment setup
   * Start parallel work
3. **Throughout:**
   * Track progress daily
   * Update sprint board
   * Communicate blockers immediately
   * Demo early and often

**Ready to start? Let’s go! 🚀**

**Document Version:** 1.0 **Last Updated:** 2025-10-06 **Owner:** Scrum Master **Approved By:** [Product Owner signature]

*Realistic 4-week plan for agile team of 5*