This file compiles the backend research and presents a clear model of the processing flow. It follows the agreed path: inspector uploads 360 videos to the dashboard, the backend runs OCR and queues work, files move into Google Drive, the sponsor PC agent pulls them into Agisoft, the model returns to Google Drive, and the backend updates the dashboard with progress and a viewer link.

Backend Flow Model

1. Actor: Inspector
   Upload .360 file to Dashboard (authenticated, resumable)

2. System: Dashboard Frontend → Backend API
   Backend receives chunks → finalizes upload → creates Job → enqueues tasks

3. System: Workers (Celery + Redis) + OCR
   Extract first frames → OCR whiteboard ID → set confidence → rename canonical
4. Build manifest.json → push video + manifest to Google Drive /ToProcess/{ID}

5. System: Google Drive (Shared Drive)
   Holds ToProcess inputs and later Processed outputs

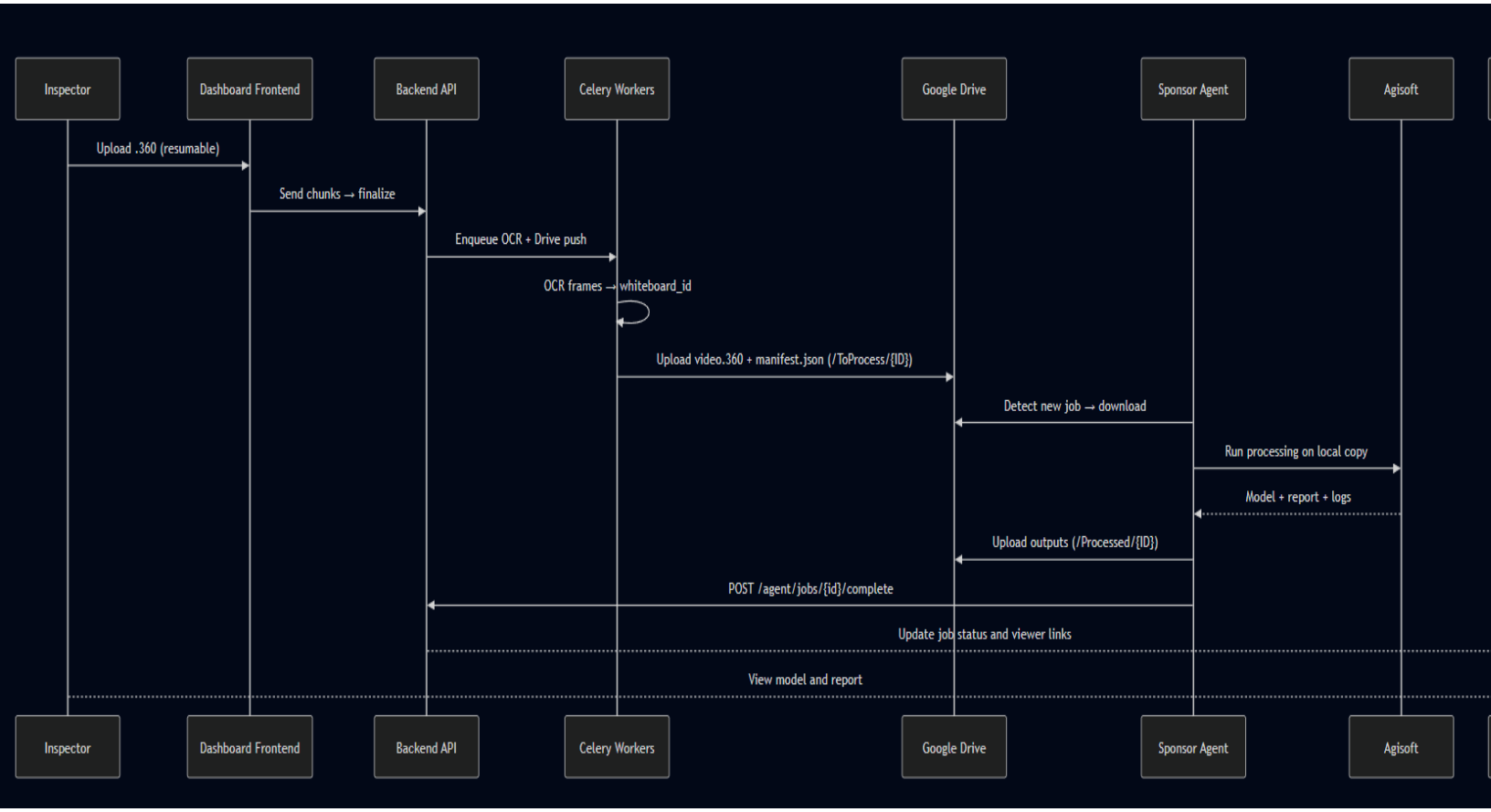
6. Actor: Sponsor PC Agent
   Detect new job on Drive → verify size and checksum → pull to local → run Agisoft

7. System: Agisoft Metashape
   Produce 3D model + report + logs

8. System: Sponsor PC Agent → Google Drive
   Upload 3D model and report to /Processed/{ID}

9. System: Backend Recon
   Detect Processed artifacts (agent callback or Drive Changes poll)
10. Update job status and progress → store Drive file IDs

11. Actor: Reviewer/Inspector on Dashboard
    Open job → view model and report in the dashboard



Detailed Components

Upload and Ingest

- Resumable upload to handle large .360 files on unstable networks. Chunks tracked by upload_id and byte ranges.

- On finalize, backend creates a video record and a job in status UPLOADED. A Celery task is enqueued for OCR.
- Basic file checks: extension whitelist, MIME sniffing, max size, optional SHA-256 for dedupe.

Queue and Workers (Celery + Redis)

- Celery runs three main tasks: run_ocr, push_to_drive, and reconcile_results.
- Retries use exponential backoff. Idempotency keys prevent double processing.
- Scheduler runs periodic scans for Drive Changes and cleanup tasks.

OCR and Canonical Naming

- Worker grabs first N frames, runs the Python OCR, and returns {whiteboard_id, confidence, crops[]}.
- If confidence is below threshold, the job moves to NEEDS_REVIEW. Otherwise it moves to OCR_OK.
- Files are renamed to a canonical form like WHITEBOARDID_YYYYMMDD.360.

Google Drive Integration

- Service Account writes to a Shared Drive that the sponsor can access.
- Folder contract:

  /ToProcess/{WHITEBOARDID}/

  video.360
  manifest.json
  /Processed/{WHITEBOARDID}/
  model/* (e.g., GLB or OBJ)
  report/{WHITEBOARDID}_report.pdf
  logs/*
  done.json

- manifest.json includes job id, checksum, file size, creation time, and OCR confidence.
- Sponsor may pay for more storage. Files can live in Drive per sponsor.

Sponsor Agent and Agisoft

- A small Windows service watches Drive for new manifest.json and a stable video.360.
- The agent downloads to the local disk, verifies checksum, then runs an Agisoft pipeline through the Python API.
- On success, the agent uploads the model, report, and logs to /Processed/{ID} and posts a completion callback to the backend.
- On failure, it writes error.json and calls a failure endpoint with a log reference.

Results Reconciliation and Progress

- The backend updates states in this order: UPLOADED → OCR_OK/NEEDS_REVIEW → PUSHED_TO_DRIVE → PROCESSING → PROCESSED or FAILED.
- Progress updates flow to the frontend over WebSockets, so inspectors see live status.
- The backend stores Drive file IDs for all artifacts and exposes short-lived links or proxied downloads.

Dashboard Viewing

- The dashboard lists jobs and shows status, timestamps, and any notes.
- On open, the viewer loads the model from Drive and renders it in a Three.js or <model-viewer> component.
- The report can be previewed or downloaded from Drive.

Operations and Risk Notes

- **Throughput**: Celery concurrency tuned to expected daily volume.
- **Quotas**: Drive API limits and storage usage monitored with alerts.
- **Resilience**: retries for transient errors and a dead-letter queue for stuck jobs.
- **Security**: Google OAuth for users, RBAC in the API, least privilege for the Service Account, all traffic over HTTPS.
- **Retention**: files live in Drive per sponsor plan. Cleanup policy for temp upload space and debug crops.