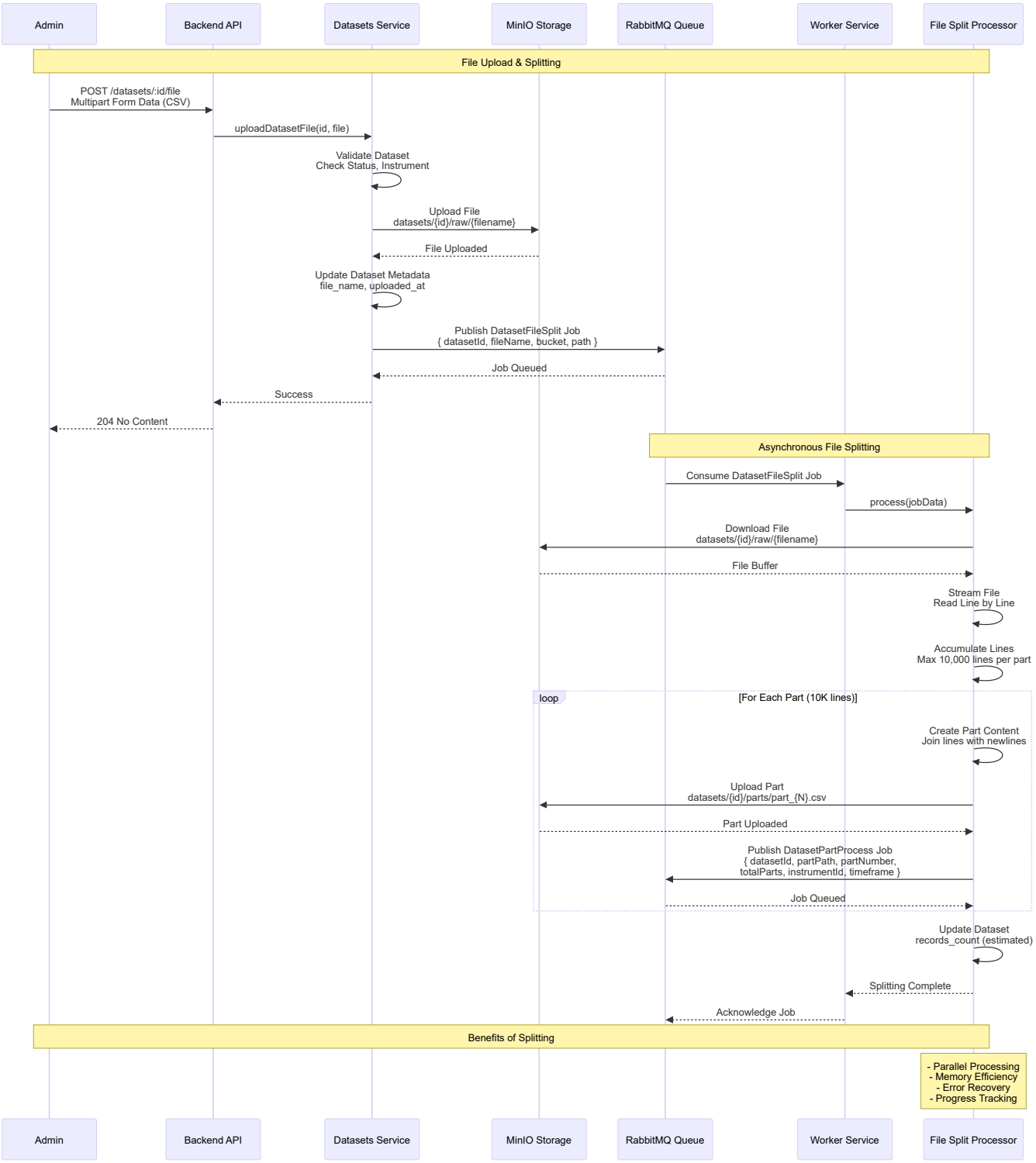
Admin
Backend API
Datasets Service
MinIO Storage
RabbitMQ Queue
Worker Service
File Split Processor
File Upload & Splitting
POST /datasets/:id/file
Multipart Form Data (CSV)
uploadDatasetFile(id, file)
Validate Dataset
Check Status, Instrument
Upload File
datasets/{id}/raw/{filename}
File Uploaded
Update Dataset Metadata
file_name, uploaded_at
Publish DatasetFileSplit Job
{ datasetId, fileName, bucket, path }
Job Queued
Success
204 No Content
Asynchronous File Splitting
Consume DatasetFileSplit Job
process(jobData)
Download File
datasets/{id}/raw/{filename}
File Buffer
Stream File
Read Line by Line
Accumulate Lines
Max 10,000 lines per part
loop
[For Each Part (10K lines)]
Create Part Content
Join lines with newlines
Upload Part
datasets/{id}/parts/part_{N}.csv
Part Uploaded
Publish DatasetPartProcess Job
{ datasetId, partPath, partNumber,
totalParts, instrumentId, timeframe }
Job Queued
Update Dataset
records_count (estimated)
Splitting Complete
Acknowledge Job
Benefits of Splitting
- Parallel Processing
- Memory Efficiency
- Error Recovery
- Progress Tracking
Admin
Backend API
Datasets Service
MinIO Storage
RabbitMQ Queue
Worker Service
File Split Processor