

Day 2 — Document → Chunks

Goal

Convert raw documents into clean, well-sized chunks suitable for retrieval.

Chunking Principles

Chunks must be semantically coherent.

Chunks must be large enough for meaning, small enough for retrieval.

Flow

Raw Document → Paragraph Split → Merge/Trim → Chunk Objects

Why This Matters

Bad chunking leads to irrelevant retrieval and downstream hallucinations.