**Rough Estimation**

1. **Estimate Document Length in Tokens**
   * A **typical document** (1-2 pages) has **500-1,500 words** → **~750-2,250 tokens**.
   * A **longer PDF** (e.g., research paper) might have **5,000+ tokens**.
   * Very Long PDFs (~10,000 tokens each, 1,000 PDFs)
2. **Summarization Token Cost**
   * If each summary is **~10-20%** of the original document size:
     + A 1,000-token document → **100-200 tokens per summary**
     + A 5,000-token document → **500-1,000 tokens per summary**
3. **Total Tokens for 1,000 PDFs**
   * If each document is **1,500 tokens** and summarized into **~200 tokens**:
     + **Input tokens:** 1,500 × 1,000 = **1.5M tokens**
     + **Output tokens:** 200 × 1,000 = **200K tokens**
     + **Total tokens:** **1.7M tokens**
     + **Total tokens used**: **5.5M - 6M tokens**
     + **Total tokens used on very long documents**: **11M - 12M tokens**
   * Long PDFs (~5,000 tokens each, 1,000 PDFs)
     + **Input tokens:** 5,000 × 1,000 = 5M tokens
     + **Output tokens (10-20% summary size):** ~500-1,000 × 1,000 = 0.5M - 1M tokens
     + **Total tokens used:** 5.5M - 6M tokens
   * Very Long PDFs (~10,000 tokens each, 1,000 PDFs)
     + **Input tokens**: 10,000 × 1,000 = 10M tokens
     + **Output tokens**: 1,000-2,000 × 1,000 = 1M - 2M tokens
     + **Total tokens used**: 11M - 12M tokens