## **1. Storage Format**

* **PDF Files**
  + Store as **raw PDFs** (.pdf format).
  + Set **Content-Type:** application/pdf.
  + Enable **server-side encryption (SSE-S3, SSE-KMS)** for sensitive data.
* **Text Files**
  + Store as **plain text** (.txt), **JSON** (.json), or **CSV** (.csv), depending on the data structure.
  + Set **Content-Type:**
    - text/plain (for .txt)
    - application/json (for .json)
    - text/csv (for .csv)
  + Use **UTF-8 encoding** to handle special characters.

## **2. Compression (Optional)**

* **Gzip (.gz)** or **Zstandard (.zst)** can help reduce storage costs, especially for large text files.
* PDFs are already compressed, so compression may not help much there.

## **3. Naming Conventions (S3 Key Structure)**

Use structured folder paths for easy retrieval:  
bash  
CopyEdit  
my-bucket/

scanned-docs/2024/01/invoice\_123.pdf

scanned-docs/2024/01/report\_456.pdf

text-files/2024/01/logs/server-log.txt

text-files/2024/01/data\_789.json

Use meaningful names with timestamps (YYYYMMDD) if versioning is needed.  
bash  
CopyEdit  
scanned-docs/invoice\_123\_20240129.pdf

text-files/report\_456\_20240129.txt

## **4. Metadata & Tags**

Add metadata for easy retrieval:  
json  
CopyEdit  
{

"Content-Type": "application/pdf",

"x-amz-meta-document-type": "invoice",

"x-amz-meta-user-id": "456"

}

* Use **S3 object tagging** to categorize files (e.g., "type": "scanned-document").

## **5. Storage & Retrieval Optimization**

* **Standard S3 Storage** for frequently accessed files.
* **S3 Intelligent-Tiering** for cost efficiency.
* **S3 Glacier** for long-term archival (if retrieval speed is not an issue).