[Cơ Bản] 3 . Tối ưu hóa Dung Lượng Lưu Trữ với Nén Tập tin.

**PHÂN TÍCH HIỆU QUẢ TIẾT KIỆM DUNG LƯỢNG THEO LOẠI TẬP TIN**

**1. Tập tin văn bản (TXT, DOCX, CSV, JSON, XML, v.v.)**

| **Định dạng gốc** | | | **Tỷ lệ nén ước tính** | | | **Nhận xét** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| .txt, .csv, .xml, .json | | | | 80-95% | | | Nén rất tốt, vì chứa nhiều dữ liệu trùng lặp, ký tự có cấu trúc dễ nén. | | |
| .docx,.xlsx | | 30-50% | | | Đã được nén sẵn (theo định dạng ZIP), nên công cụ nén thêm chỉ cải thiện nhẹ. | | |

**Kết luận**: Các tập tin văn bản thuần túy (TXT, CSV) nén cực tốt với ZIP/RAR/7z.  
- **Định dạng nén hiệu quả**: ZIP, 7z.

**2. Tập tin hình ảnh (JPG, PNG, BMP, TIFF, v.v.)**

| **Định dạng gốc** | **Tỷ lệ nén ước tính** | **Nhận xét** |
| --- | --- | --- |
| .bmp, .tiff | 70-90% | Không nén hoặc nén kém, dễ nén lại bằng ZIP. |
| .png | 10-30% | Đã là lossless nén, khó nén thêm nhiều. |
| .jpg, .webp | 0-5% | Đã nén theo chuẩn lossy, gần như không giảm thêm. |

📌 **Kết luận**:

* Ảnh không nén (BMP, TIFF) có thể nén rất tốt.
* Ảnh đã nén (JPG, PNG) khó cải thiện thêm.  
  ✅ **Định dạng nén hiệu quả**: JPG/WebP cho ảnh chia sẻ; ZIP cho ảnh BMP/TIFF gốc.

**3. Tập tin video (MP4, AVI, MKV, MOV, v.v.)**

| **Định dạng gốc** | **Tỷ lệ nén ước tính** | **Nhận xét** |
| --- | --- | --- |
| .avi, .mov (raw) | 40-70% | Video chưa nén có thể giảm mạnh dung lượng. |
| .mp4, .mkv (H.264, H.265) | 0-10% | Đã được nén rất hiệu quả bằng thuật toán lossy. |

**Kết luận**:

* Video chưa nén nén tốt.
* Video nén sẵn (MP4, MKV) hầu như không thể nén thêm bằng ZIP.
* **Định dạng nén hiệu quả**: H.265 (HEVC), AV1 cho nén video; ZIP/RAR không phù hợp.

**4. Tập tin âm thanh (WAV, MP3, FLAC, AAC)**

| **Định dạng gốc** | **Tỷ lệ nén ước tính** | **Nhận xét** |
| --- | --- | --- |
| .wav | 50-70% | Âm thanh thô, nén rất tốt. |
| .mp3, .aac, .ogg | 0-5% | Đã nén (lossy), gần như không thể nén thêm. |
| .flac | 10-20% | Nén không mất dữ liệu, có thể giảm chút. |

**Kết luận**:

* Âm thanh chưa nén (WAV) có thể giảm dung lượng rõ rệt.
* Âm thanh đã nén (MP3, AAC) thì không thể nén thêm.  
   **Định dạng nén hiệu quả**: MP3, AAC cho chia sẻ; ZIP dùng với WAV.

**II. TẠI SAO MỘT SỐ TẬP TIN KHÔNG THỂ NÉN NHIỀU?**

**1. Đã được nén sẵn**

* Ví dụ: JPG, MP4, MP3, PNG đều sử dụng thuật toán nén riêng.
* Khi nén lại bằng công cụ như ZIP, kết quả gần như không đổi (hoặc thậm chí **lớn hơn** nếu thêm metadata).

**2. Không có nhiều dữ liệu lặp lại**

* Các thuật toán nén hoạt động tốt khi có **mẫu dữ liệu trùng lặp**.
* File nhị phân (ví dụ: mã hóa, mã máy) hoặc đã mã hóa thường không có mẫu lặp, khiến việc nén kém hiệu quả.

**3. Dữ liệu ngẫu nhiên hoặc đã qua xử lý entropy**

* Các file sinh ra từ thuật toán mã hóa hoặc sinh số ngẫu nhiên thường có **entropy cao**, rất khó nén.

TỔNG KẾT: ĐỊNH DẠNG NÉN TỐT NHẤT THEO LOẠI TẬP TIN

| **Loại tập tin** | **Tốt nhất nên dùng** | **Lý do** |
| --- | --- | --- |
| Văn bản thuần | .zip, .7z | Tỷ lệ nén cao (lên đến 95%). |
| Hình ảnh chưa nén | .zip, .7z | BMP, TIFF nén tốt. |
| Hình ảnh nén (JPG, PNG) | Không cần nén thêm | Không tiết kiệm được nhiều. |
| Video đã nén (MP4, MKV) | Giữ nguyên định dạng | Nén thêm không hiệu quả. |
| Video thô (AVI, MOV) | Chuyển sang H.264/H.265 | Nén mạnh, giảm hàng chục GB. |
| Âm thanh thô (WAV) | .zip, hoặc chuyển MP3 | Nén tốt hoặc chuyển đổi định dạng. |
|  |  |  |

# **Thực hành** qua ảnh .