

# Gestión Avanzada de Conjuntos de Datos

Descarga, Procesamiento y Aplicación en Proyectos Reales

---

---

# Parte I: Ecosistema de Formatos

Fundamentos técnicos y estrategias de adquisición de datos modernos.

---

# Formatos: Estructurado vs Semi-est.



## CSV (Commas)

Formato tabular clásico. Ideal para datos masivos estructurados con esquemas fijos.

Es altamente eficiente para lectura secuencial y procesamiento incremental.



## JSON & XML

Soporte para jerarquías y datos anidados en estructuras de árbol.

JSON es el estándar de facto para APIs modernas debido a su ligereza y facilidad de manipulación en Python/JS.

# Estrategias de Adquisición



## APIs REST

Extracción directa desde servidores mediante protocolos HTTP. Entrega usualmente formatos JSON/XML en tiempo real.



## Web Scraping

Captura de datos no estructurados desde HTML para convertirlos en estructuras procesables (CSV/JSON).



## Datasets

Uso de repositorios masivos como Kaggle o datos gubernamentales abiertos en múltiples extensiones.

# Herramientas de Programación

El dominio de bibliotecas específicas es clave para la eficiencia en ingeniería de datos:



## Pandas

El estándar para la manipulación y análisis de tablas de datos. Permite la lectura eficiente de formatos estructurados mediante `read_csv` y `read_json`.



## Requests

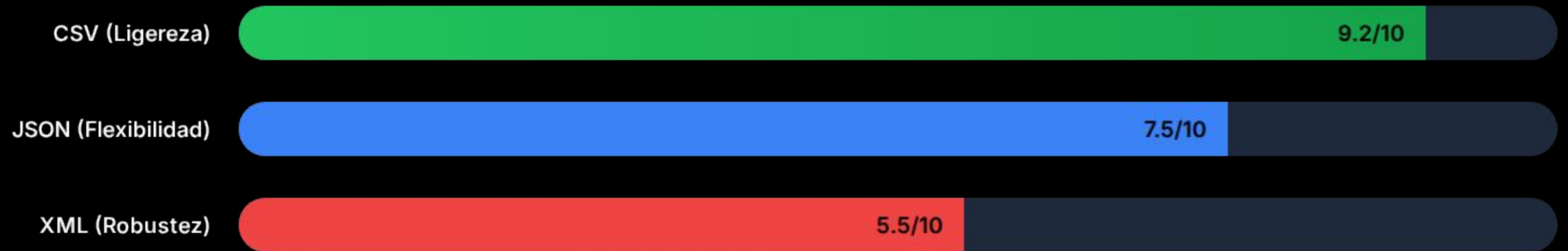
Librería fundamental para la gestión de descargas HTTP y la interacción simplificada con APIs REST y servicios web externos.



## ElementTree

Herramienta nativa de alto rendimiento diseñada para el procesamiento, parsing y estructuración jerárquica de archivos en formato XML.

# Eficiencia: **Tamaño** vs Velocidad



*Comparativa basada en tiempos de carga y volumen de metadatos redundantes.*

---

# Parte II: **Aplicación** Práctica

Casos de estudio y flujos de trabajo en proyectos de ingeniería de datos.

---

# Bibliotecas y Herramientas

Formato	Biblioteca Python	Uso Principal	Complejidad
CSV	Pandas / csv	Análisis Tabular	Baja
JSON	json / requests	Intercambio Web	Media
XML	lxml / ElementTree	Configuración/Legados	Alta
SQL	SQLAlchemy	Persistencia	Media

# Impacto del Procesamiento

80%

Tiempo Ahorrado en ETL

## Automatización de Pipelines

La implementación de scripts de lectura y escritura automática reduce drásticamente el error humano y permite el procesamiento de terabytes de información en minutos, optimizando la toma de decisiones empresariales.

- ✓ **Eficiencia Operativa:** Eliminación de flujos manuales repetitivos.
- ✓ **Procesamiento Veloz:** Análisis de terabytes en cuestión de minutos.
- ✓ **Decisión Instantánea:** Información clave lista para el negocio.

# Casos de Estudio Reales



E-commerce

Gestión de catálogos masivos mediante JSON para precios dinámicos.



Streaming

Algoritmos de recomendación basados en metadatos de usuario.



Meteorología

Integración de sensores remotos mediante protocolos XML complejos.

# Flujo de Trabajo en Plenaria

01

## Preparación de Datos

Selección de fuentes confiables y limpieza inicial (Data Wrangling).

02

## Demostración de Código

Exposición de scripts de lectura/escritura optimizados.

03

## Análisis de Resultados

Interpretación técnica de la eficiencia del formato elegido.

04

## Discusión de Casos

Debate sobre la aplicabilidad del método en entornos de producción real.

# ¿Preguntas o Consultas?

Sesión de Plenaria y Retroalimentación Técnica

DRA. OLGA LOPEZ FORTIZ

---

# Image Sources



[http://www.contractoruk.com/sites/default/files/2026-06/shutterstock\\_2682443667.jpg](http://www.contractoruk.com/sites/default/files/2026-06/shutterstock_2682443667.jpg)

Source: [www.contractoruk.com](http://www.contractoruk.com)

---



<https://slidebazaar.com/wp-content/uploads/2025/06/digital-marketing-kpis-dashboard-template-powerpoint-google-slides.jpg>

Source: [slidebazaar.com](https://slidebazaar.com)

---



[https://miro.medium.com/1\\*Lz2QGFkqvYSKE3yIGxB-A.png](https://miro.medium.com/1*Lz2QGFkqvYSKE3yIGxB-A.png)

Source: [uxmag.medium.com](https://miro.medium.com)

---



<https://projects.arduinocontent.cc/cover-images/f8042e0f-36cf-452b-9dd3-49b8a93a5750.jpg>

Source: [projecthub.arduino.cc](https://projects.arduinocontent.cc)

---