



**UNIVERSIDAD CATÓLICA**  
de Colombia

# Uso de Herramientas Tecnológicas para Data Analytics y Big Data

**Diego Alberto Rincón Yáñez MSc**  
**darincon@ucatolica.edu.co**

Afiliada a la Federación Internacional de Universidades Católicas (FIUC)  
w w w . u c a t o l i c a . e d u . c o

# Agenda



UNIVERSIDAD CATÓLICA  
de Colombia

- Legislación
  - Transparencia y del Derecho de Acceso a la Información Pública.
  - Protección de Datos Personales.
- Tipos de Archivos Abiertos.
  - Formatos
    - Plano
    - Binario (Documentos e Imagen)

# Ley de Transparencia



UNIVERSIDAD CATÓLICA  
de Colombia

- **Nombre :** LEY 1712 DE 2014
- **Creada en:** Sentencia c-274/13 – [en línea](#)  
– 9 de mayo de 2013
- **Vigencia desde:** Diario Oficial No. 49.084 de 6 de marzo de 2014 – [en línea](#)

# Principios



UNIVERSIDAD CATÓLICA  
de Colombia

- **Transparencia.**
- **Buena fe.**
- **Facilitación**
- **No discriminación.**
- **Gratuidad.**
- **Celeridad.**
- **Eficacia.**
- **Calidad de la información.**
- **Divulgación proactiva de la información.**
- **Responsabilidad en el uso de la información.**

# Definiciones



UNIVERSIDAD CATÓLICA  
de Colombia

- Información.
- Información pública.
- Información pública clasificada.
- Información pública reservada.
- Publicar o divulgar.
- Sujetos obligados.
- Gestión documental.
- Documento de archivo.
- Archivo.
- Datos Abiertos.
- Documento en construcción.

# ¿A Quien Aplica?



UNIVERSIDAD CATÓLICA  
de Colombia

- Toda entidad pública.
- Los órganos, organismos y entidades estatales independientes o autónomos y de control.
- Las personas naturales y jurídicas, públicas o privadas, que presten función pública.
- Cualquier persona natural, jurídica o dependencia de persona jurídica que desempeñe función pública o de autoridad pública.
- Los partidos o movimientos políticos y los grupos significativos de ciudadanos.
- Las entidades que administren instituciones parafiscales, fondos o recursos de naturaleza u origen público.



# Ley de Protección de Datos

- **Nombre :** Ley 1581 de 2012
- **Creada en:** Sentencia C-748 de 2011– [en línea](#)  
– 6 de octubre de 2011
- **Vigencia desde:** Decreto número 1317 de 2013, de 26 de Junio de 2013 – [en línea](#)



# ¡TallerNo.1!

- ✓ Capitulo 1: 3 Personas
- ✓ Capitulo 2: 6 Personas
- ✓ Capitulo 3: 4 Personas
- ✓ Capitulo 4: 2 Personas
- ✓ Capitulo 5: 2 Personas
- ✓ Capitulo 6: 2 Personas
- ✓ Grupo Sorpresa: 5 Personas





# Tipos de Archivos Abiertos

- ¿Qué es?
- ¿Cómo Identificarlo?
- ¿Cómo Consultarlo?
- Formatos
  - Plano - ¿Cómo identificarlo?
  - Binario - ¿Cómo identificarlo?
- Métodos de Consulta



# Tipos de Formato Plano y Sintaxis

- Texto Plano – Texto normal

- CSV
  - nombre,edad
  - juan,20
  - pedro80

nombre	Edad
juan	20
pedro	80



# Tipos de Formato Plano y Sintaxis

- XML (Extensible Markup Language)

- Basado en Etiquetas

- TODA etiqueta que se abre, debe cerrarse

- Ejemplo:

```
<?xml version="1.0" encode="ISO-8859-1"?>
```

```
<agenda>
```

```
<contacto>
```

```
    <nombre>Diego Rincón</nombre> <telefono>645552388</telefono>
```

```
    <email>diego@gmail.com</email>
```

```
</contacto>
```

```
</agenda>
```



# Tipos de Formato Plano y Sintaxis

- JSON (JavaScript Object Notation)
  - Formato para el intercambios de datos
  - Alternativa a XML
  - Tipos de Datos
    - Un **número** (entero o float)
    - Un **string** (entre comillas simples)
    - Un **booleano** (true o false)
    - Un **array** (entre corchetes [] )
    - Un **objeto** (entre llaves {})
    - **Null**

# Ejemplo



UNIVERSIDAD CATÓLICA  
de Colombia

- Ejemplo:

- Fruta:

- 10 manzanas
    - 20 Peras
    - 30 Naranjas

"Nombre" : "Diego Rincón"

```
{ "NombreFruta":"Manzana" ,  
  "Cantidad":20 }
```

- Verduras

- 80 lechugas
    - 15 tomates
    - 50 pepinos

```
{  "Frutas": [  
  { "NombreFruta":"Manzana" ,  
    "cantidad":10 },  
  { "NombreFruta":"Pera" ,  
    "cantidad":20 },  
  { "NombreFruta":"Naranja" ,  
    "cantidad":30 }    ]}
```

# ¡Ejercicio No.2!



UNIVERSIDAD CATÓLICA  
de Colombia

- Ejemplo:
  - Fruta:
    - 10 manzanas
    - 20 Peras
    - 30 Naranjas
  - Verduras
    - 80 lechugas
    - 15 tomates
    - 50 pepinos

## Completa el Ejemplo

¡Valídalo!

[Validador en Línea](#)

[Visualizador JSON](#)





# Tipos de Formato Binario

- Imagen
  - Jpg
  - Gif
  - Png
  - [ppm](#)
- Documentos
  - Doc
  - Xls
  - Docx
  - PDF
- Audio
  - Mp3
  - Wav
  - Raw
  - midi
- Video
  - Mp4
  - Avi
  - mkv

[Editor Hexadecimal Portable](#)

# Tipos de Formato Binario



UNIVERSIDAD CATÓLICA  
de Colombia

- Ejemplo leer ppm desde C
- Leer PDF desde Java
  - [Apache PDF BOX](#)
- Es bien largo, mejor descarguémoslo [aquí](#)

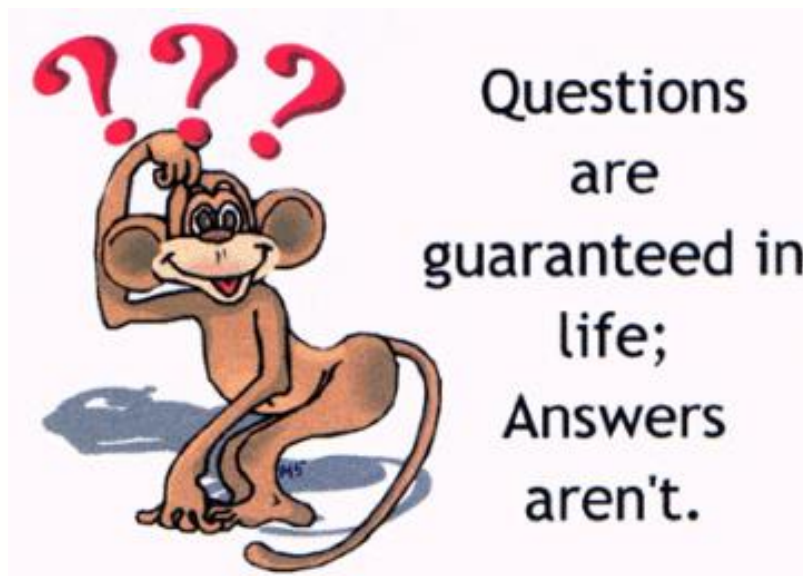
```
try {
    PDDocument document = null;
    document = PDDocument.load(new File("test.pdf"));
    document.getClass();
    if (!document.isEncrypted()) {
        PDTextStripperByArea stripper = new
        PDTextStripperByArea();
        stripper.setSortByPosition(true);
        PDTextStripper Tstripper = new
        PDTextStripper();
        String st = Tstripper.getText(document);
        System.out.println("Text:" + st);
    }
} catch (Exception e) { e.printStackTrace(); }
```





# Métodos de Consulta

- HTTP
- WSDL
- ¿otros?



# ¿Preguntas?

Diego Alberto Rincón Yáñez MCSc.  
Twitter: @d1egoprogram