
Definición arquitectónica

HAGA CLIC PARA AGREGAR SUBTÍTULO

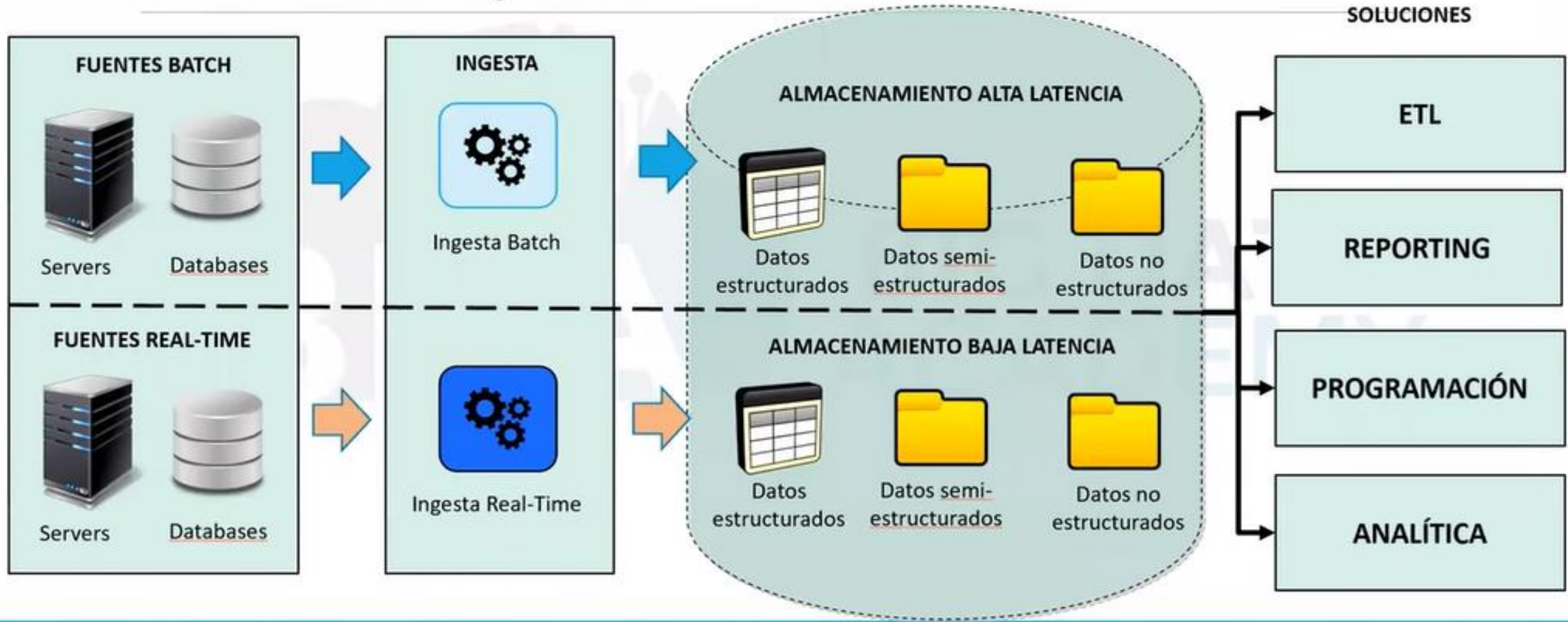
Resumen de construcción

1. **Vista conceptual [x]**
2. Vista tecnológica
3. **Vista de gobierno [x]**
4. **Vista de modelamiento [x]**
5. **Vista de calidad [x]**
6. **Vista de integración [x]**
7. **Vista de infraestructura [x]**
8. Vista de seguridad
9. Vista de DevOps
10. **Vista de patrones de diseño [x]**
11. Vista de arquetipos de código

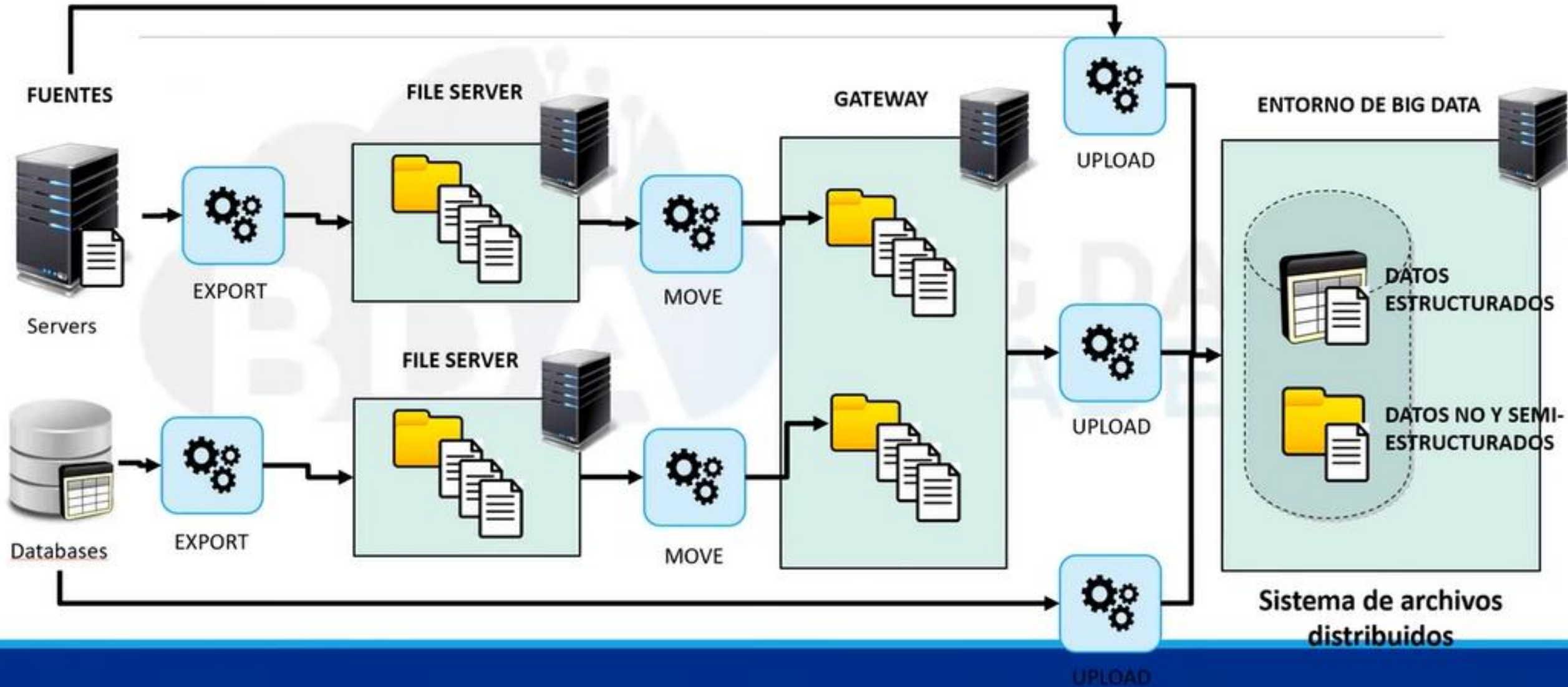
Vista conceptual

HAGA CLIC PARA AGREGAR TEXTO

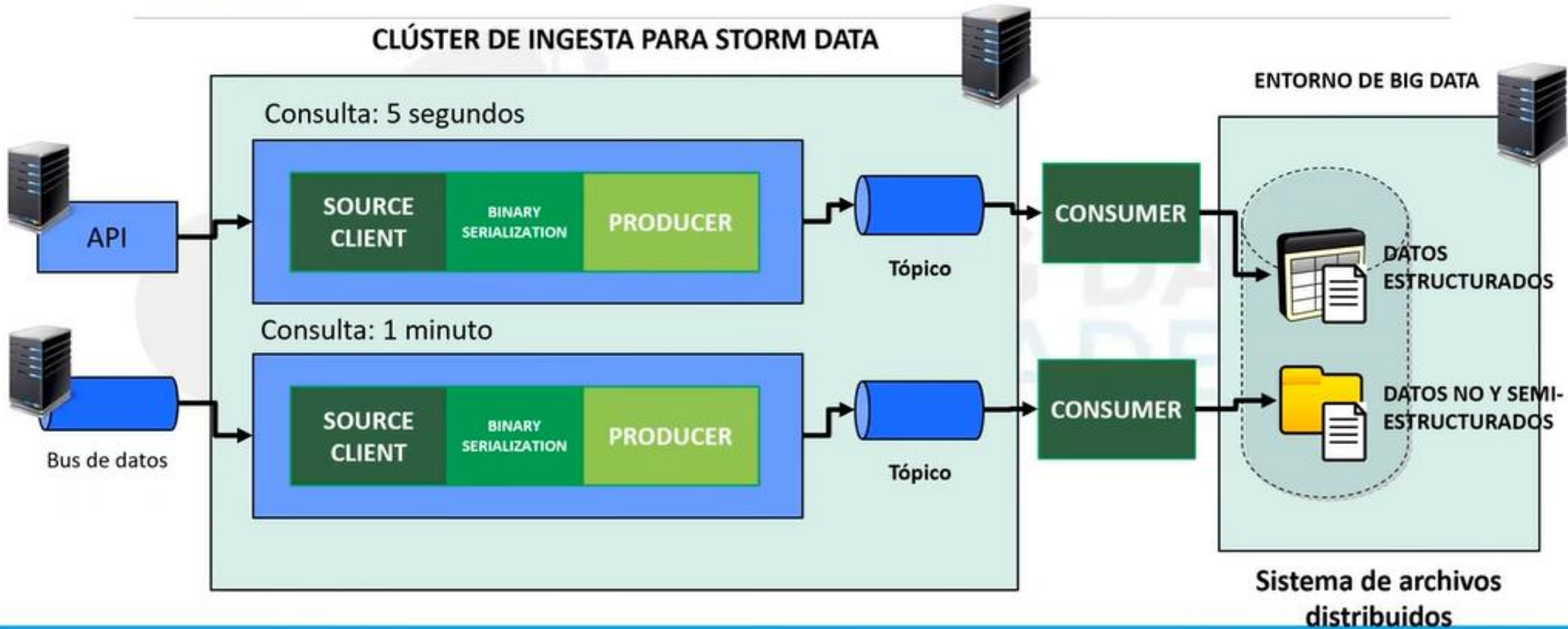
Vista conceptual



Estrategia de ingesta de fuentes BATCH

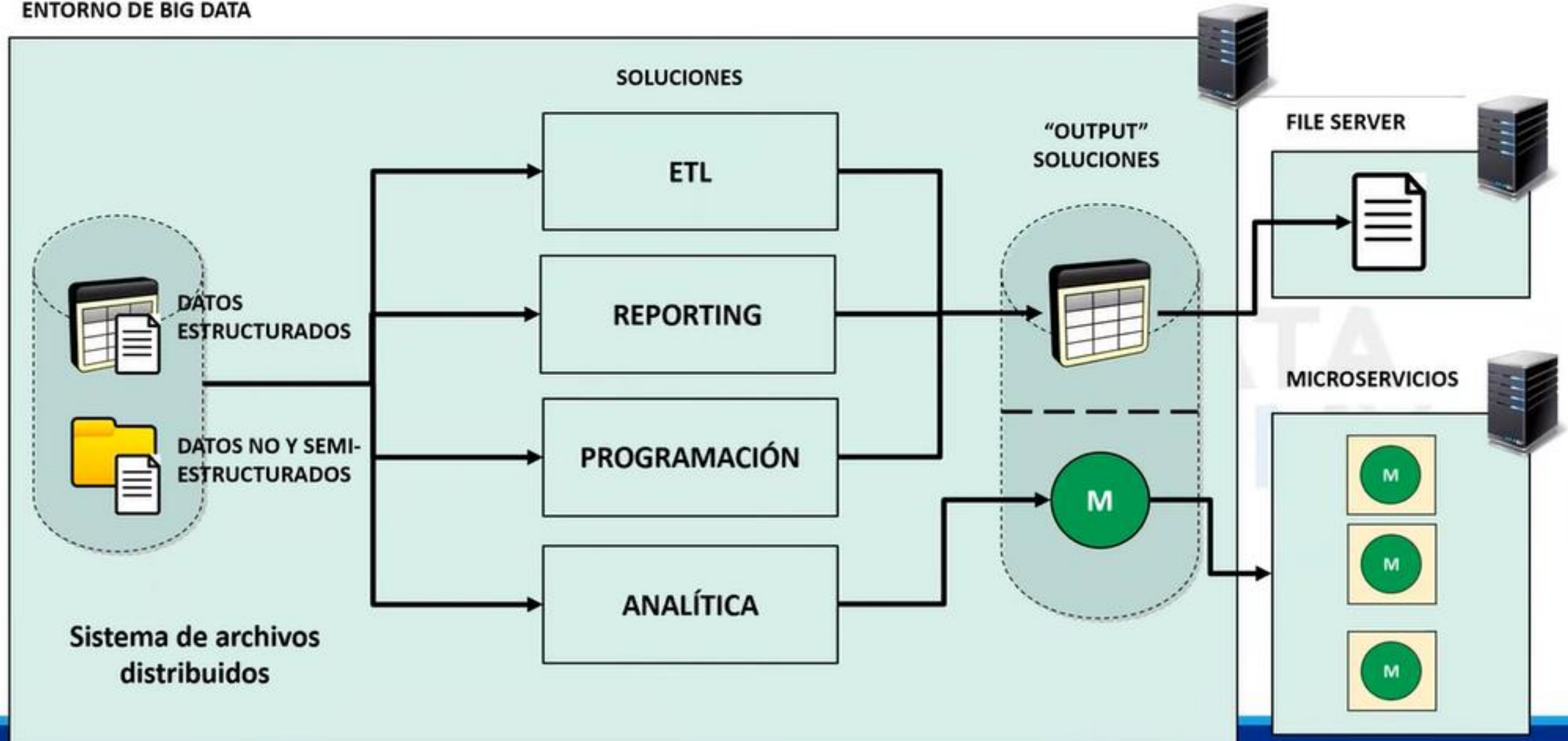


Estrategia de ingesta de fuentes REAL TIME

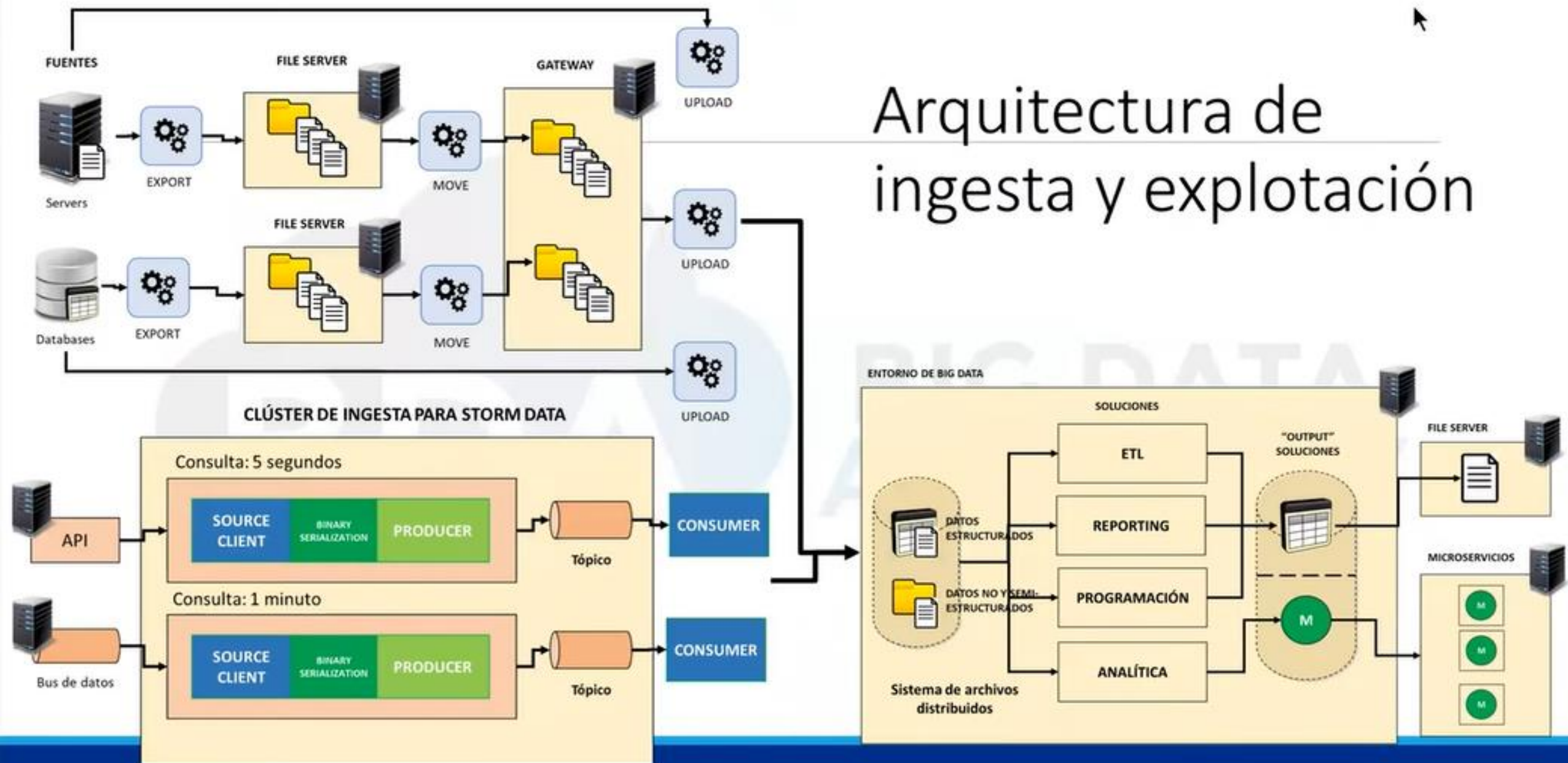


Estrategia de explotación e integración

ENTORNO DE BIG DATA



Arquitectura de ingesta y explotación



Listar fuentes de datos por ingestar

Fuentes estructuradas

BASE DE DATOS	TABLA	PERIODICIDAD
COMPRAS	PERSONA	DIARIA
COMPRAS	EMPRESA	DIARIA
COMPRAS	TRANSACCION	DIARIA
COMPRAS	PRODUCTO	SEMANAL
TARJETAS	TIPO_TARJETA	ANUAL
TARJETAS	TARJETA_CLIENTE	DIARIA

Fuentes semi-estructuradas y no estructuradas

SERVER	RUTA	TIPO	PERIODICIDAD
192.168.1.107	/data/dataset/transacciones.json	JSON	DIARIA
192.168.1.109	/files/prd/bd/compras.xml	XML	DIARIA
192.168.1.122	/datasets/user/devoluciones.csv	CSV	SEMANAL
192.168.1.181	/tornillos/*.jpg	JPG	DIARIA
192.168.1.111	/FILES/record/*.mp4	MP4	DIARIA
192.168.1.116	/FILES/audio/*.wav	WAV	MENSUAL



Vista de gobierno

HAGA CLIC PARA AGREGAR TEXTO

Datalake

ETL

MODELING

SOLUTIONS

LANDING_TMP

LANDING

UNIVERSAL

SMART



DATOS ESTRUCTURADOS



Move To Partition



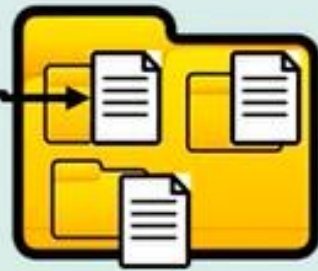
Modeling



DATOS SEMI ESTRUCTURADOS



Move To Partition



Modeling



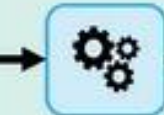
DATOS NO ESTRUCTURADOS



Move To Partition



Modeling



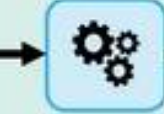
Modeling



Datamesh



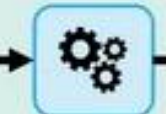
Solution



Modeling



Datamesh



Solution

M

...



Modeling



Datamesh



Solution

M

Datos estructurados, semi-estructurados y no estructurados

DATOS ESTRUCTURADOS



Tablas



Archivos con
separadores

DATOS SEMI- ESTRUCTURADOS



JSON



XML

DATOS NO ESTRUCTURADOS



IMÁGENES



VIDEOS

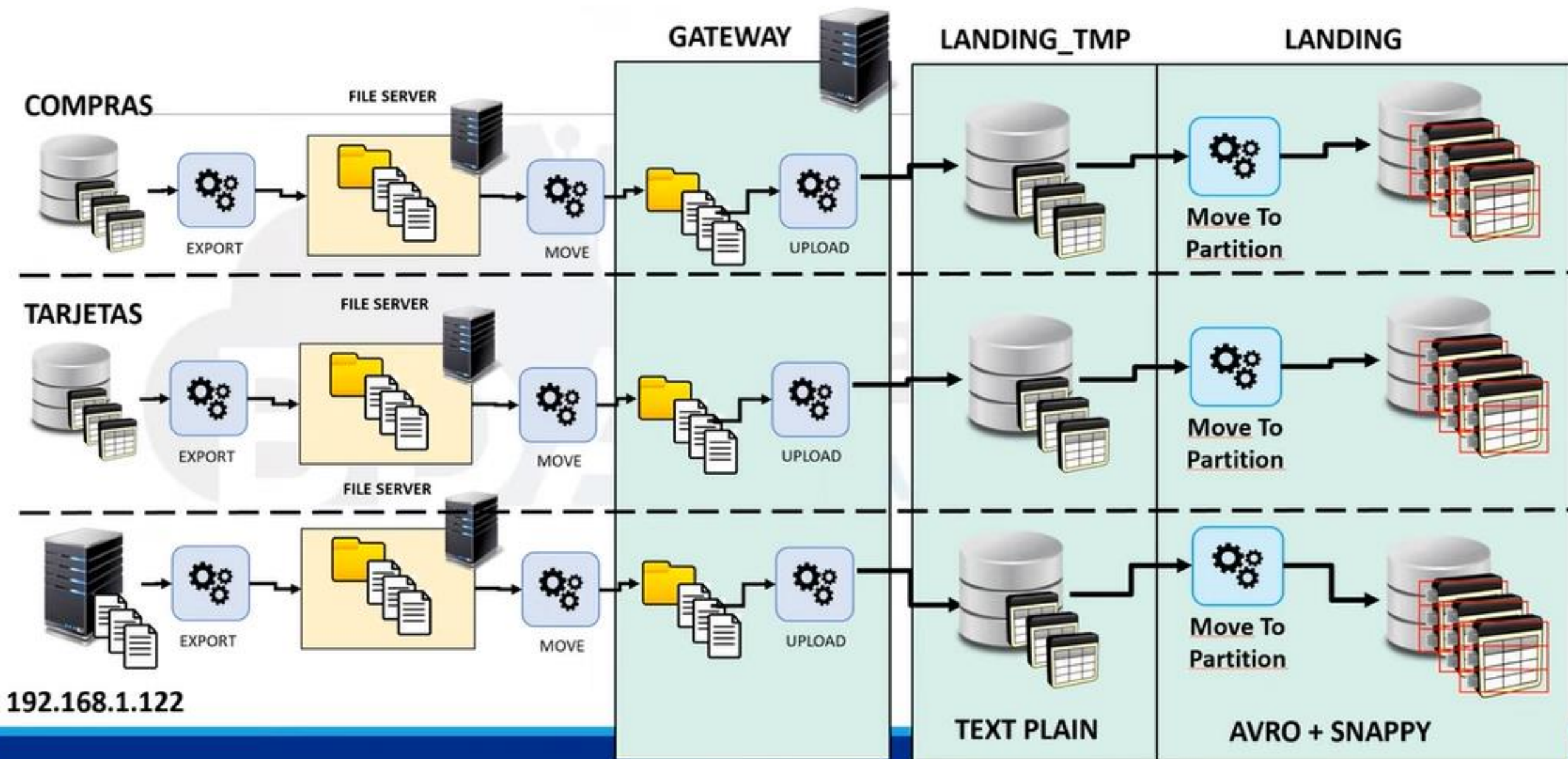


AUDIOS



DOCUMENTOS
ENRIQUECIDOS

ETL para datos estructurados



Formatos semi-estructurados

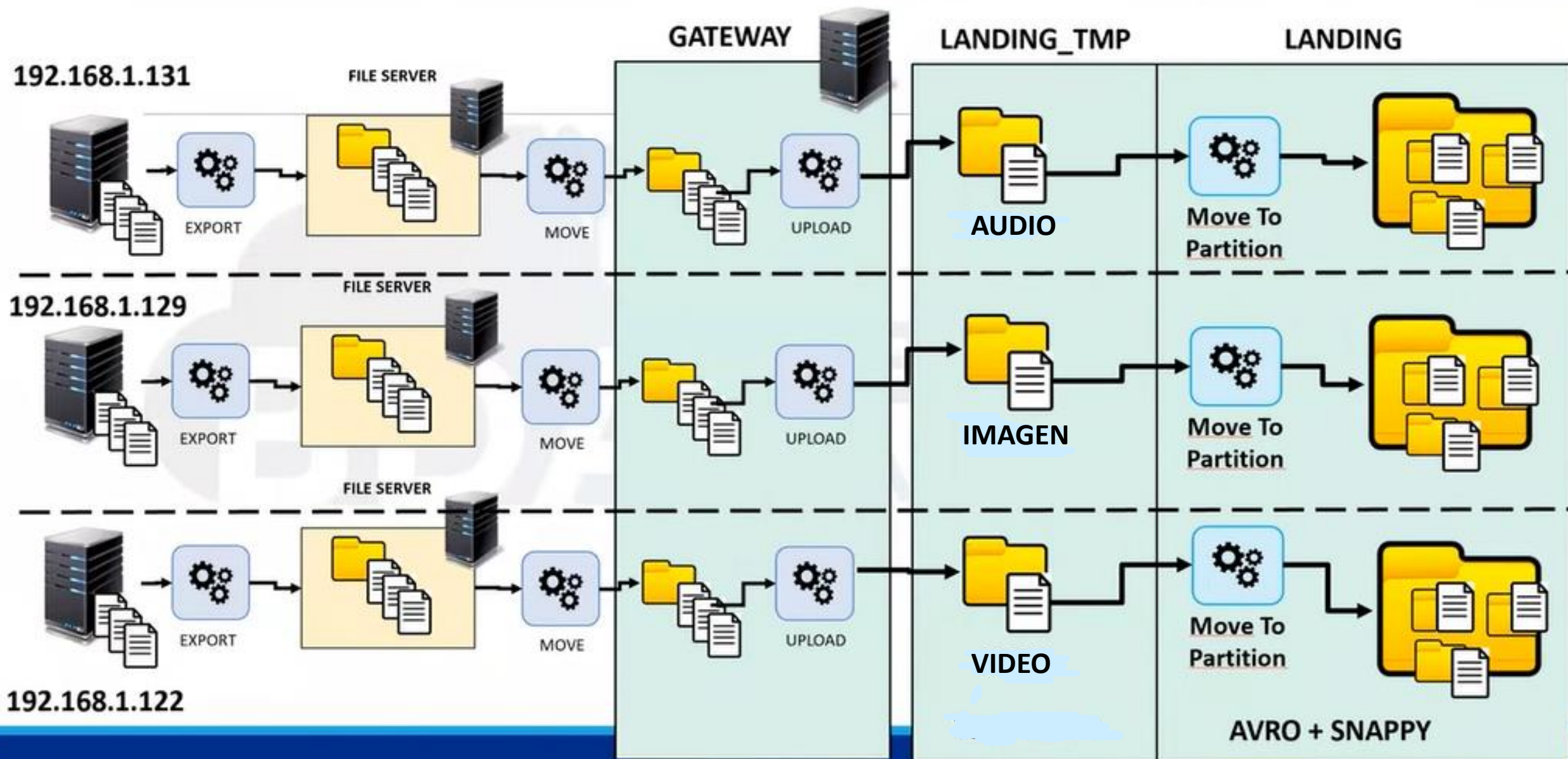
```
transacciones_complejas.json
[
  {
    "PERSONA": {
      "ID_PERSONA": 18,
      "NOMBRE": "Owen",
      "CONTACTO": [
        { "TIPO": "CORREO", "VALOR": "owen@hotmail.com" },
        { "TIPO": "CORREO", "VALOR": "owen@gmail.com" },
        { "TIPO": "TELEFONO", "VALOR": "991281120" }
      ]
    },
    "EMPRESA": {
      "ID_EMPRESA": 3,
      "NOMBRE": "Apple"
    },
    "MONTO": 1383,
    "FECHA": "2020-01-03"
  },
  {
    "PERSONA": {
      "ID_PERSONA": 30,
      "NOMBRE": "Clayton"
    },
    "EMPRESA": {
      "ID_EMPRESA": 6,
      "NOMBRE": "Google"
    },
    "MONTO": 2331,
    "FECHA": "2020-01-04"
  },
  {
    "PERSONA": {
      "ID_PERSONA": 47,
      "NOMBRE": "Morgan"
    },
    "EMPRESA": {
      "ID_EMPRESA": 3,
      "NOMBRE": "Apple"
    },
    "MONTO": 1383,
    "FECHA": "2020-01-03"
  }
]
```

ARCHIVO JSON

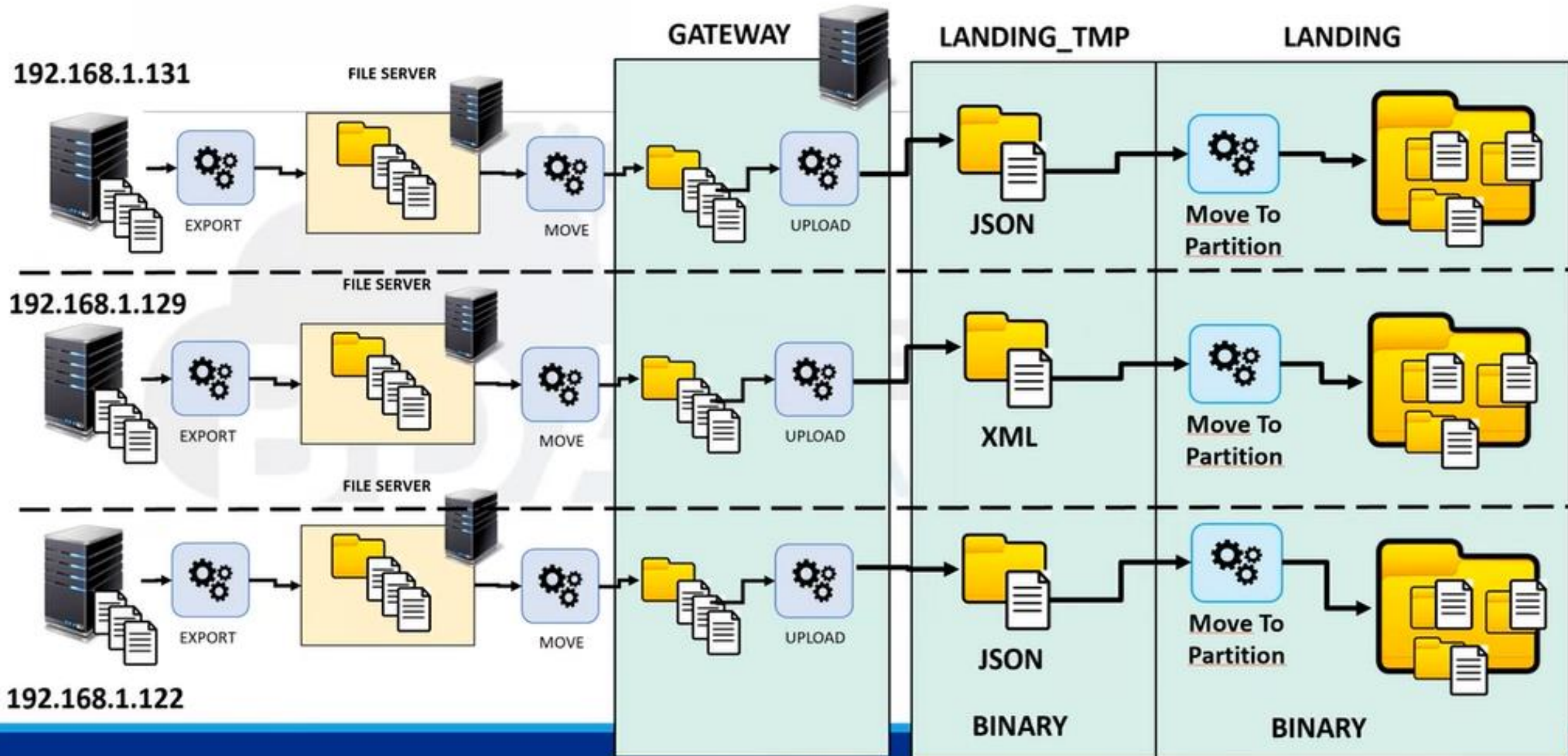
```
transacciones_complejas.xml
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <element>
    <EMPRESA>
      <ID_EMPRESA>3</ID_EMPRESA>
      <NOMBRE>Apple</NOMBRE>
    </EMPRESA>
    <FECHA>2020-01-03</FECHA>
    <MONTO>1383</MONTO>
    <PERSONA>
      <CONTACTO>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>owen@hotmail.com</VALOR>
        </element>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>owen@gmail.com</VALOR>
        </element>
        <element>
          <TIPO>TELEFONO</TIPO>
          <VALOR>991281120</VALOR>
        </element>
      </CONTACTO>
      <ID_PERSONA>18</ID_PERSONA>
      <NOMBRE>Owen</NOMBRE>
    </PERSONA>
  </element>
  <element>
    <EMPRESA>
      <ID_EMPRESA>6</ID_EMPRESA>
      <NOMBRE>Google</NOMBRE>
    </EMPRESA>
    <FECHA>2020-01-04</FECHA>
    <MONTO>2331</MONTO>
    <PERSONA>
      <CONTACTO>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>morgan@gmail.com</VALOR>
        </element>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>morgan@hotmail.com</VALOR>
        </element>
        <element>
          <TIPO>TELEFONO</TIPO>
          <VALOR>991281120</VALOR>
        </element>
      </CONTACTO>
      <ID_PERSONA>47</ID_PERSONA>
      <NOMBRE>Morgan</NOMBRE>
    </PERSONA>
  </element>
  <element>
    <EMPRESA>
      <ID_EMPRESA>3</ID_EMPRESA>
      <NOMBRE>Apple</NOMBRE>
    </EMPRESA>
    <FECHA>2020-01-03</FECHA>
    <MONTO>1383</MONTO>
    <PERSONA>
      <CONTACTO>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>owen@hotmail.com</VALOR>
        </element>
        <element>
          <TIPO>CORREO</TIPO>
          <VALOR>owen@gmail.com</VALOR>
        </element>
        <element>
          <TIPO>TELEFONO</TIPO>
          <VALOR>991281120</VALOR>
        </element>
      </CONTACTO>
      <ID_PERSONA>18</ID_PERSONA>
      <NOMBRE>Owen</NOMBRE>
    </PERSONA>
  </element>
</root>
```

ARCHIVO XML

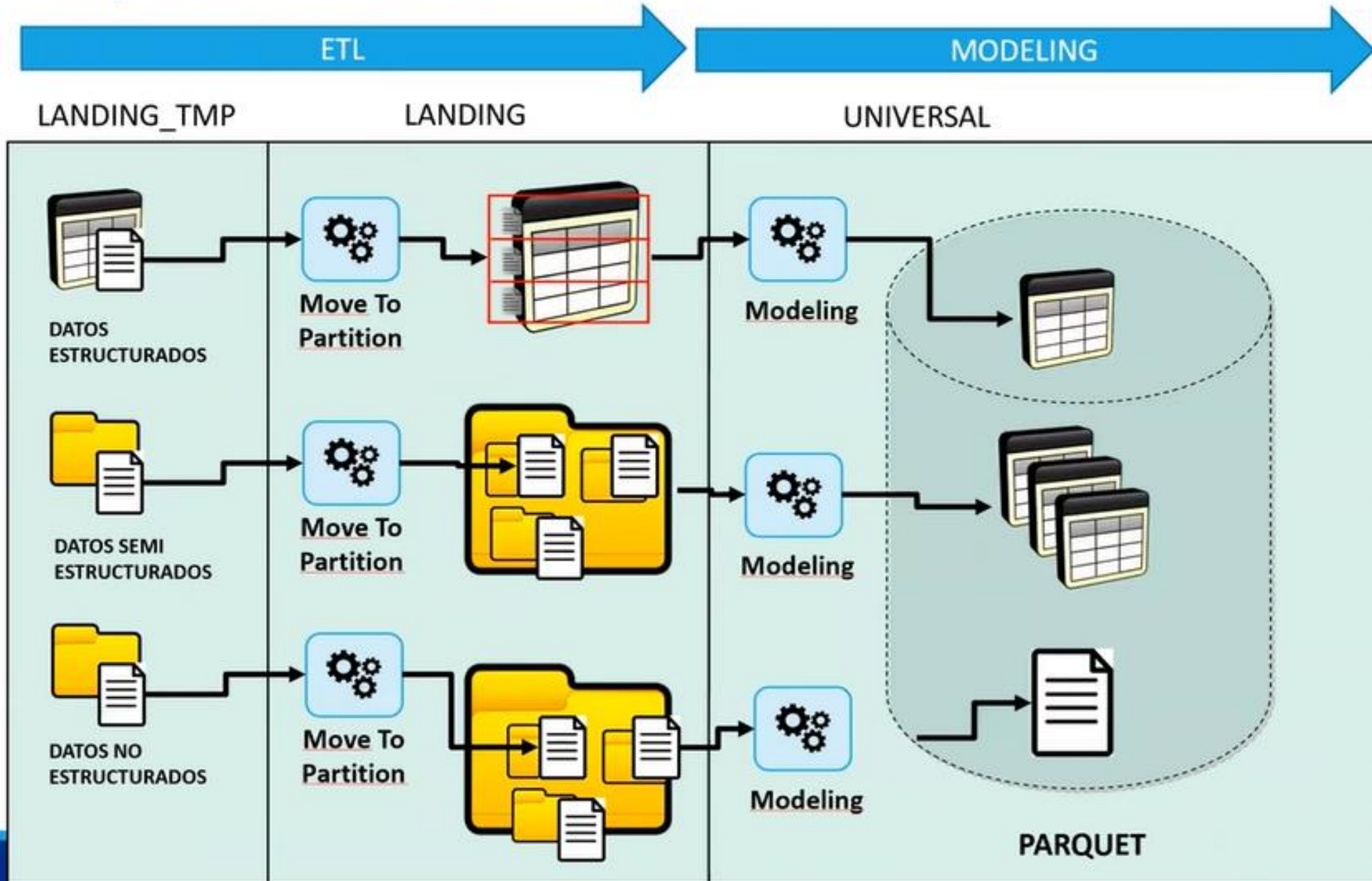
ETL para datos semiestructurados



ETL para datos no estructurados



Capa de modelamiento



Definición del modelo datos [Vista de modelamiento]

Definir para cada tabla:

- Campos
- Tipos de dato

TABLA: PERSONA

CAMPO	TIPO DE DATO
ID	CHAR(8)
NOMBRE	STRING
SEXO	STRING
FECHA_NACIMIENTO	DATE
SALARIO	DECIMAL (10,2)

TABLA: EMPRESA

CAMPO	TIPO DE DATO
ID	CHAR(8)
NOMBRE	STRING

TABLA: TRANSACCION

CAMPO	TIPO DE DATO
ID_PERSONA	CHAR(8)
ID_EMPRESA	CHAR(8)
FECHA_HORA	TIMESTAMP
MONTO	DECIMAL(10,2)

Reglas de calidad [Vista de calidad]

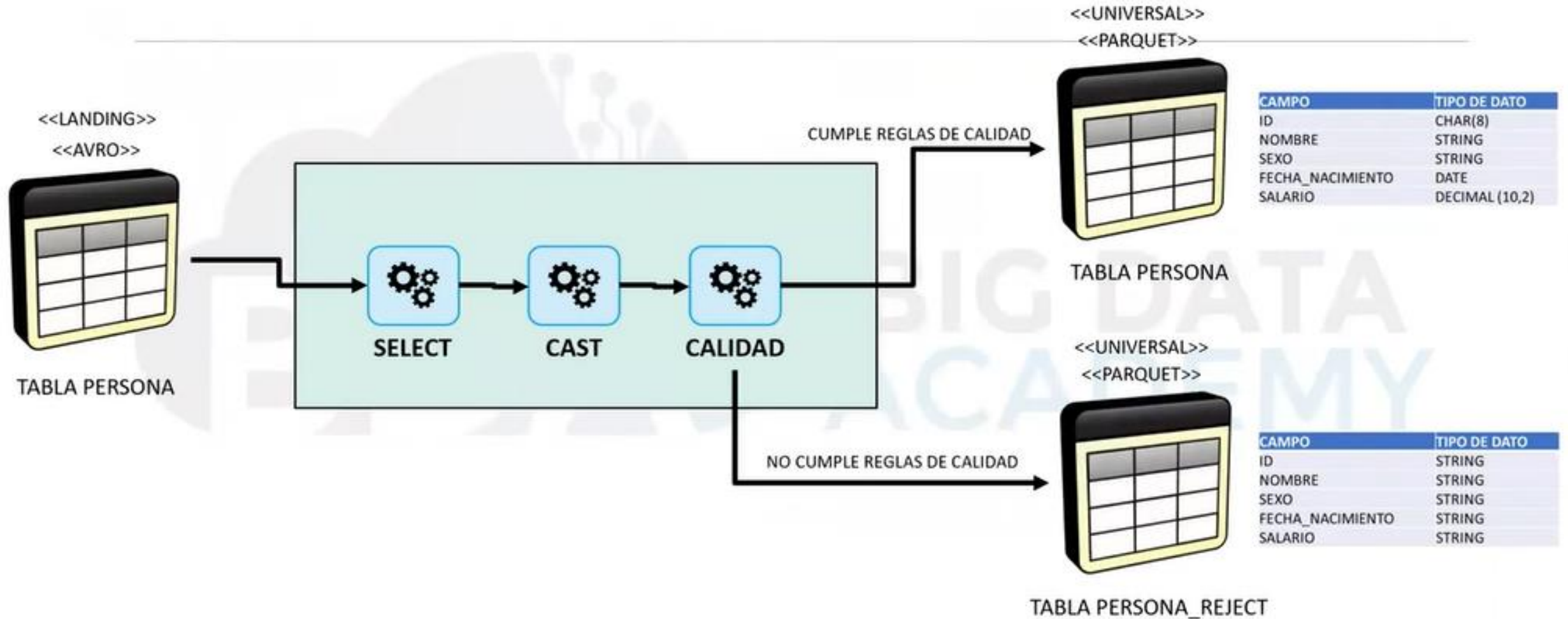
TABLA: PERSONA

CAMPO	TIPO DE DATO
ID	CHAR(8)
NOMBRE	STRING
SEXO	STRING
FECHA_NACIMIENTO	DATE
SALARIO	DECIMAL (10,2)

REGLAS DE CALIDAD:

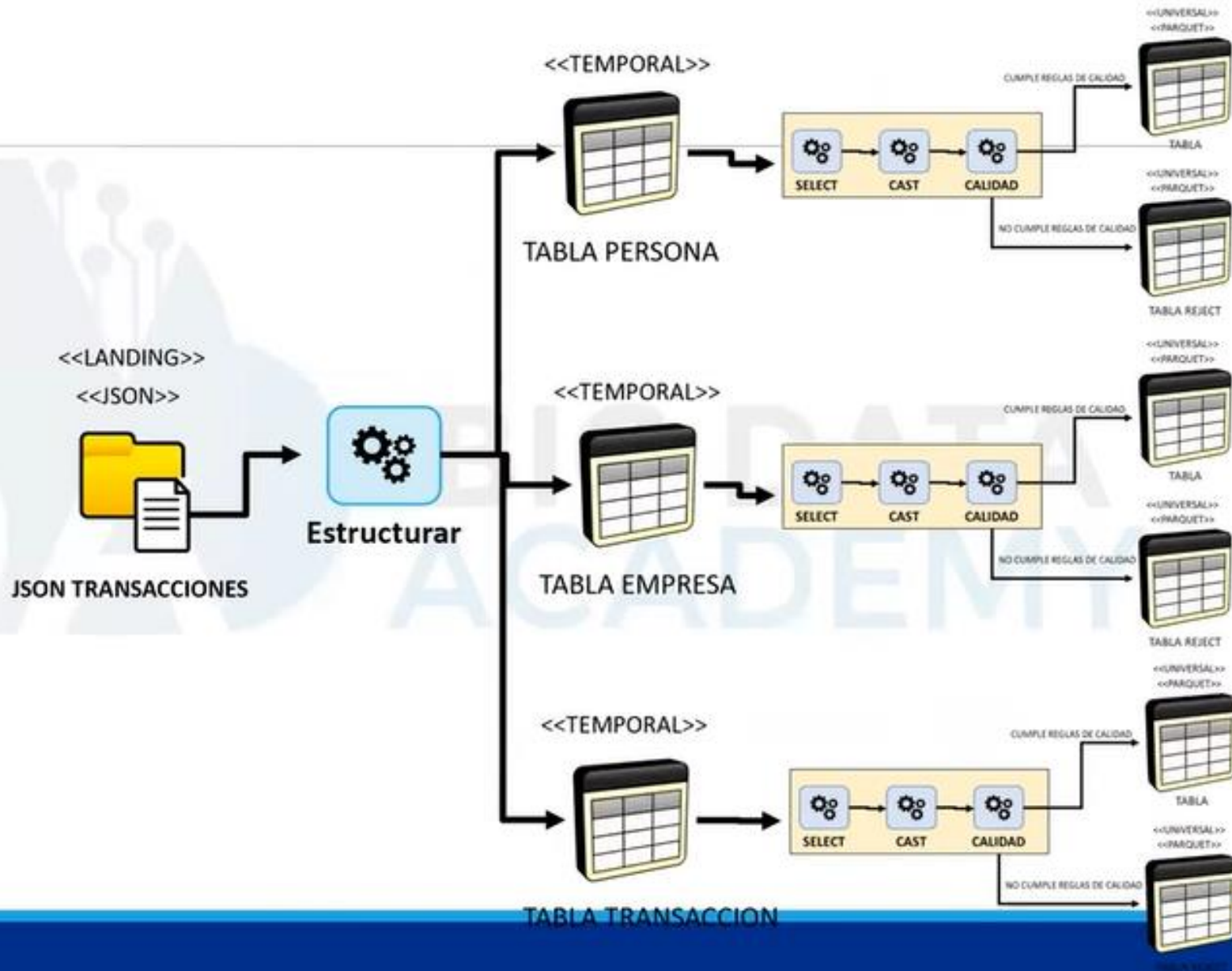
1. ID NO PUEDE SER NULO
2. SEXO PUEDE TOMAR EL VALOR DE "M", "F", O NULL
3. EL SALARIO DEBE SER MAYOR A CERO
4. EL SALARIO DEBE SER MENOR A 100000

Modelamiento de datos estructurados



Modelamiento de datos semi-estructurados

```
transacciones_complejas.json
[
  {
    "PERSONA": {
      "ID_PERSONA": 18,
      "NOMBRE": "Owen",
      "CONTACTO": [
        { "TIPO": "CORREO", "VALOR": "owen@hotmail.com" },
        { "TIPO": "CORREO", "VALOR": "owen@gmail.com" },
        { "TIPO": "TELEFONO", "VALOR": "991281120" }
      ]
    },
    "EMPRESA": {
      "ID_EMPRESA": 3,
      "NOMBRE": "Apple"
    },
    "MONTO": 1383,
    "FECHA": "2020-01-03"
  },
  {
    "PERSONA": {
      "ID_PERSONA": 30,
      "NOMBRE": "Clayton"
    },
    "EMPRESA": {
      "ID_EMPRESA": 6,
      "NOMBRE": "Google"
    },
    "MONTO": 2331,
    "FECHA": "2020-01-04"
  },
  {
    "PERSONA": {
      "ID_PERSONA": 47,
      "NOMBRE": "Clayton"
    },
    "EMPRESA": {
      "ID_EMPRESA": 6,
      "NOMBRE": "Google"
    },
    "MONTO": 2331,
    "FECHA": "2020-01-04"
  }
]
```

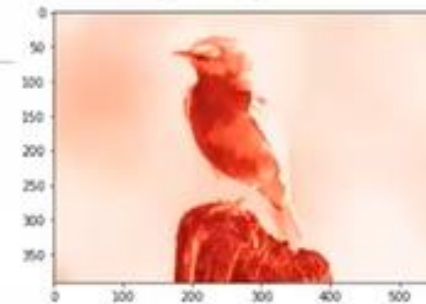


Estructurando datos no estructurados

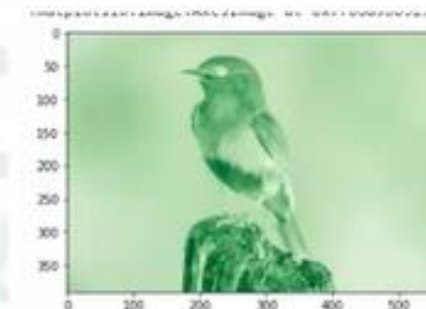
[Imágenes]



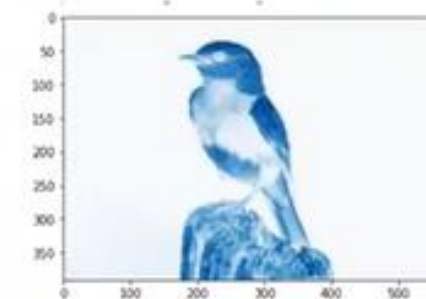
Imagen original



Canal R



Canal G

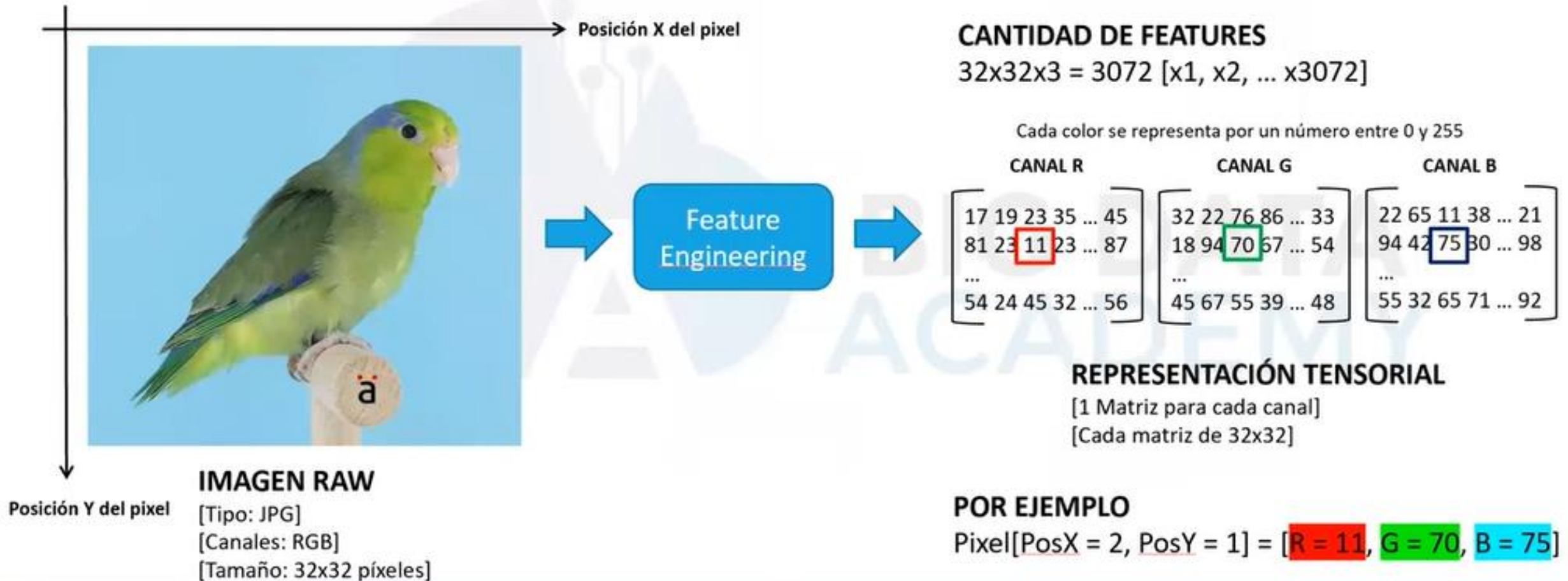


Canal B

Representación tensorial
[alto, ancho, canal]

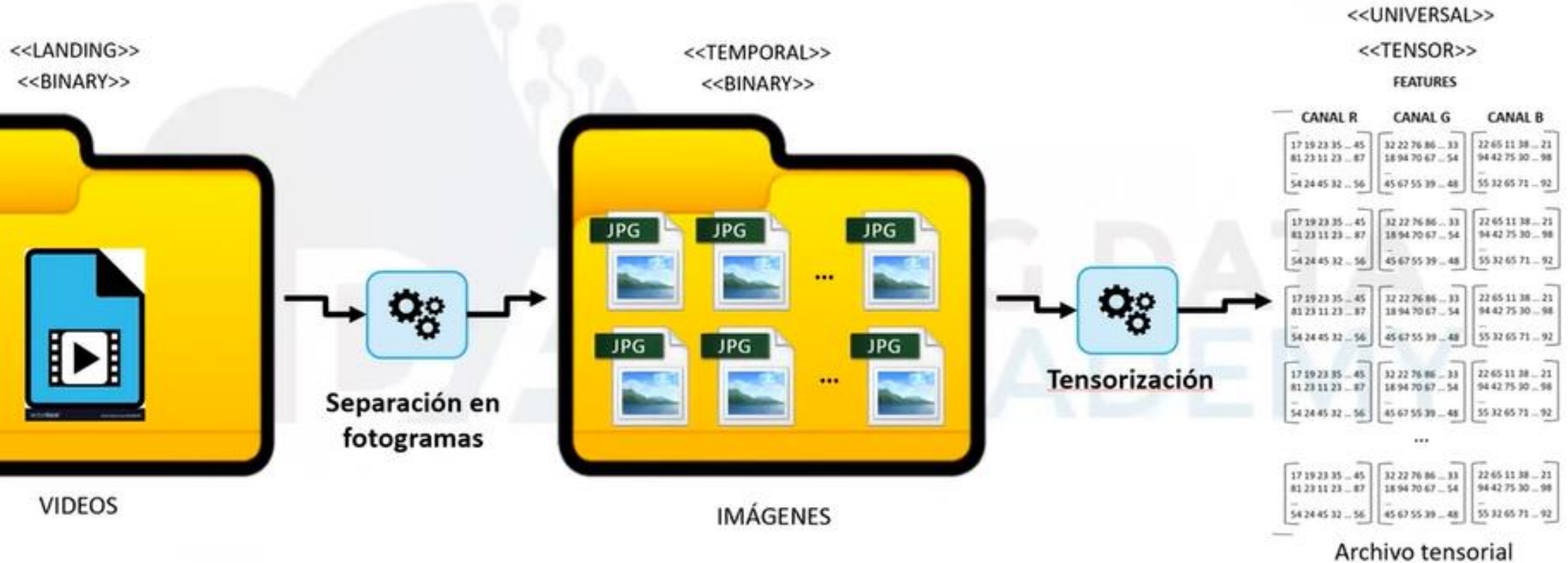
Estructurando datos no estructurados

[Imágenes]

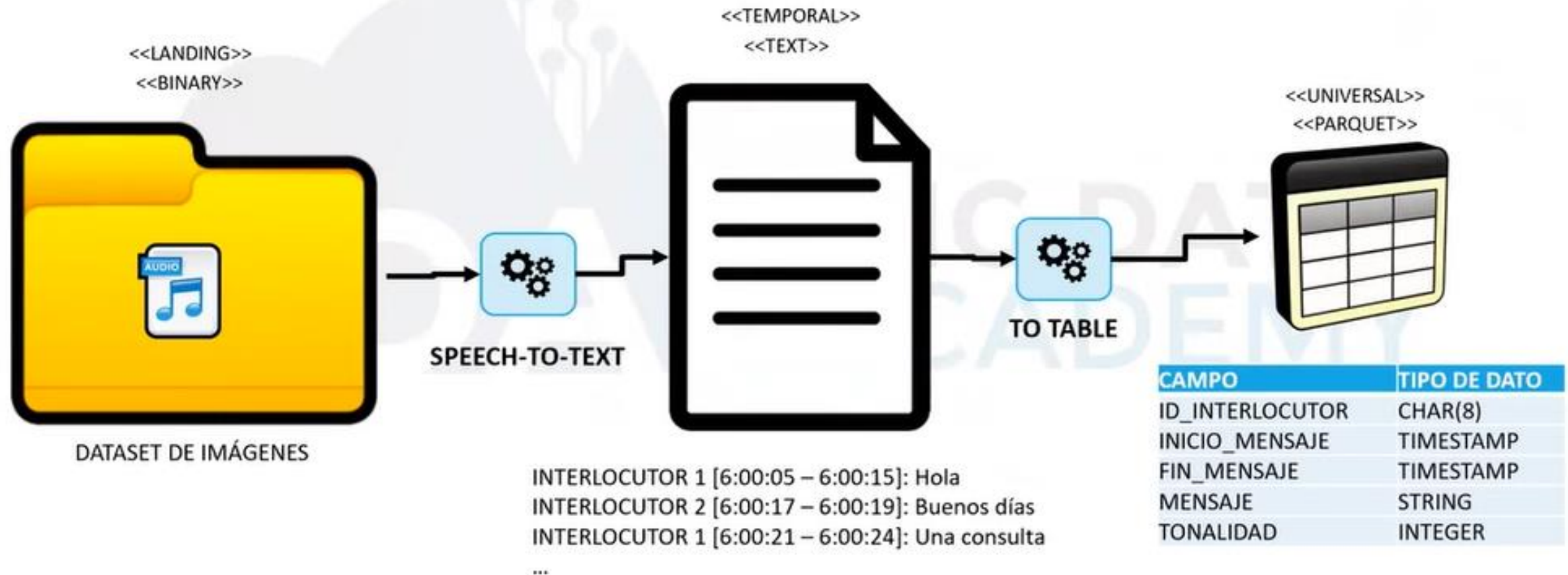




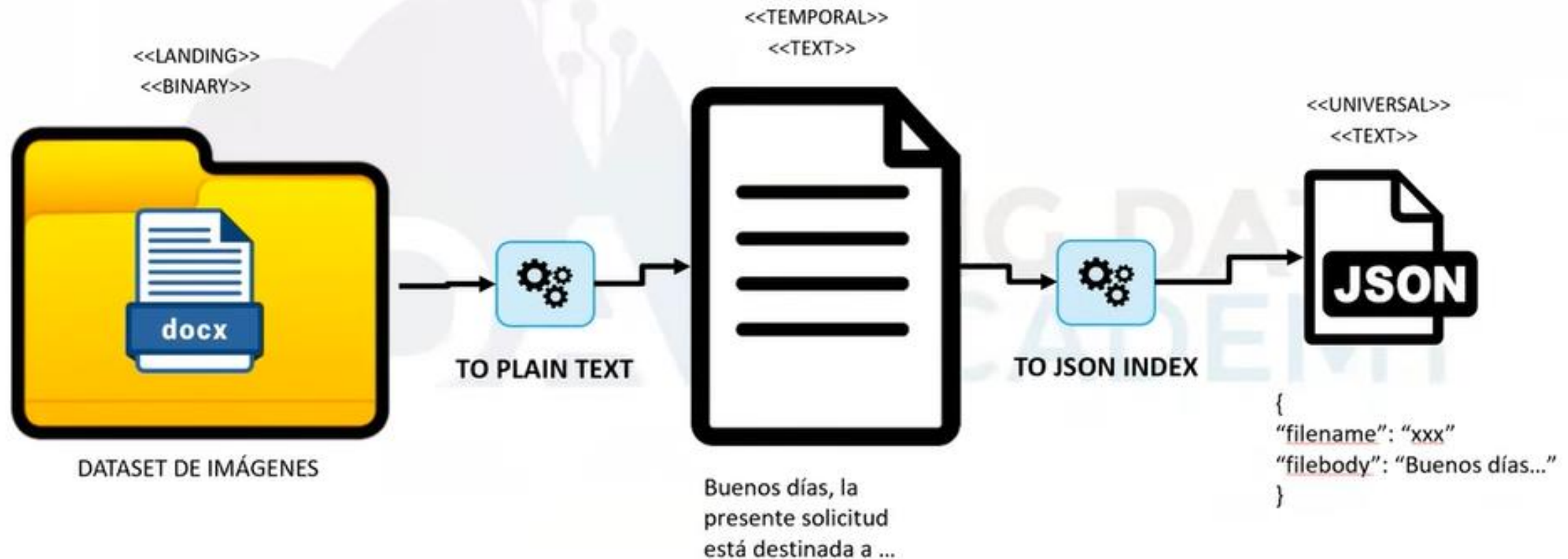
Estructurando datos no estructurados [Videos]



Estructurando datos no estructurados [Audios]



Estructurando datos no estructurados [Documentos enriquecidos]





Resumen

HAGA CLIC PARA AGREGAR TEXTO

Siguiente paso: Vista tecnológica y Programar arquetipos

Ingesta estructurada [Archivo CSV]

Ingesta semi-estructurada [Archivo JSON]

Ingesta no estructurada [Archivo JPG]

Modelamiento estructurado [Select, Cast y Calidad]

Modelamiento semi-estructurado [Estructurar]

Modelamiento no estructurado [Tensorizar]

IMPORTANTE: TENER LISTA LA CUENTA DATABRICKS

Siguiente paso: Definición de las soluciones

ENTORNO DE BIG DATA

