

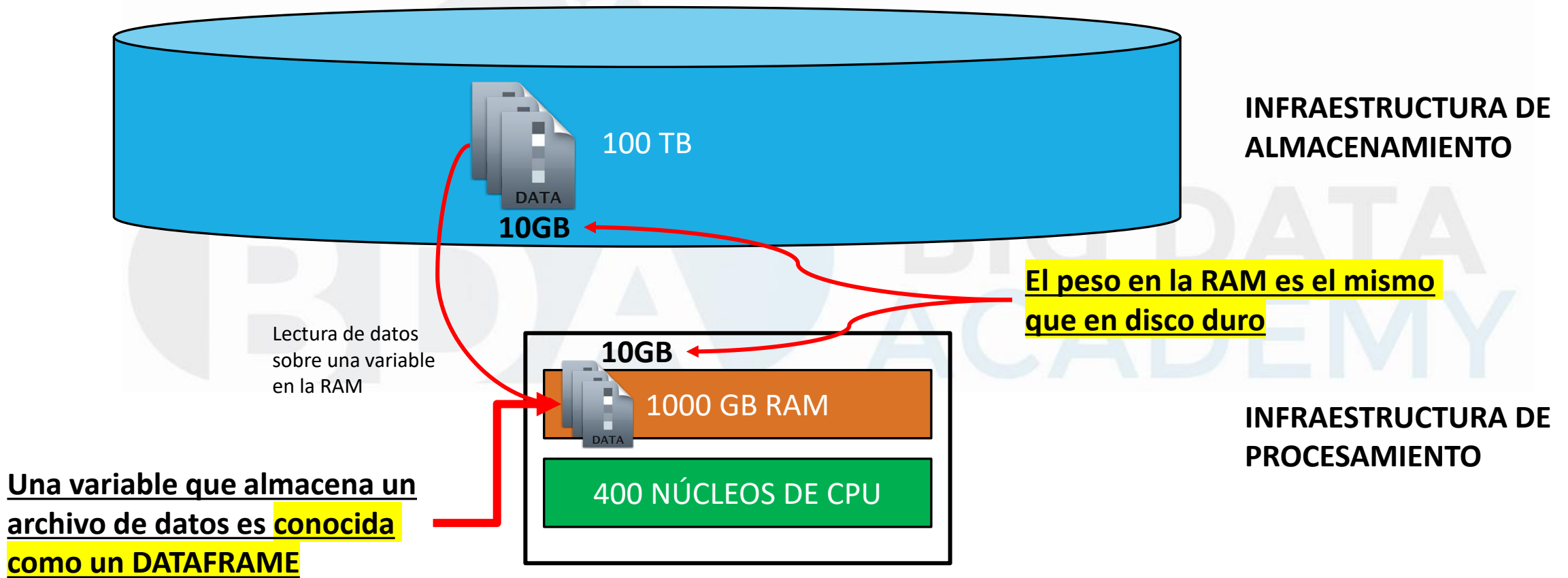
---

# Laboratorio 09

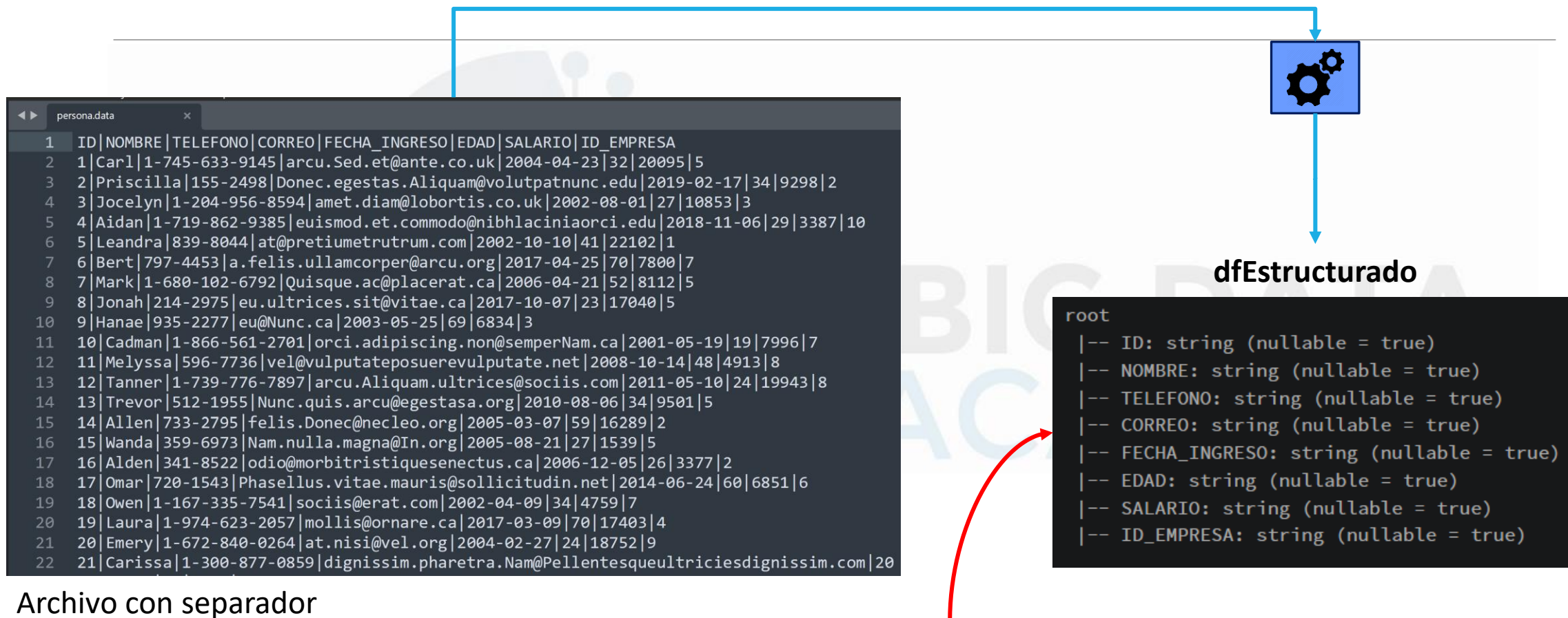
LECTURA DE DATOS

---

# Lectura y procesamiento de datos



# Dataframes estructurados desde archivos



Archivo con separador

Todos los registros tienen el mismo esquema de metadatos

# Registros semi-estructurados

```
{
  "PERSONA":{
    "ID_PERSONA":"26",
    "NOMBRE_PERSONA":"Brenden",
    "EDAD":33,
    "SALARIO":20549.0,
    "CONTACTO":[
      {"PREFIJO":"59","TELEFONO":"9811935"},
      {"PREFIJO":"53","TELEFONO":"9423163"}
    ]
  },
  "EMPRESA":{
    "ID_EMPRESA":"5",
    "NOMBRE_EMPRESA":"Amazon"
  },
  "TRANSACCION":{
    "MONTO":2628.0,
    "FECHA":"2021-01-23"
  }
}
```

Un registro semi-estructurado define su propio esquema de metadatos y puede tener subcampos



```
root
|-- EMPRESA: struct (nullable = true)
|   |-- ID_EMPRESA: string (nullable = true)
|   |-- NOMBRE_EMPRESA: string (nullable = true)
|-- PERSONA: struct (nullable = true)
|   |-- CONTACTO: array (nullable = true)
|   |   |-- element: struct (containsNull = true)
|   |   |   |-- PREFIJO: string (nullable = true)
|   |   |   |-- TELEFONO: string (nullable = true)
|   |-- EDAD: long (nullable = true)
|   |-- ID_PERSONA: string (nullable = true)
|   |-- NOMBRE_PERSONA: string (nullable = true)
|   |-- SALARIO: double (nullable = true)
|-- TRANSACCION: struct (nullable = true)
|   |-- FECHA: string (nullable = true)
|   |-- MONTO: double (nullable = true)
```

## Archivo JSON

Alonso Melgarejo [alonsoraulmgs@gmail.com]