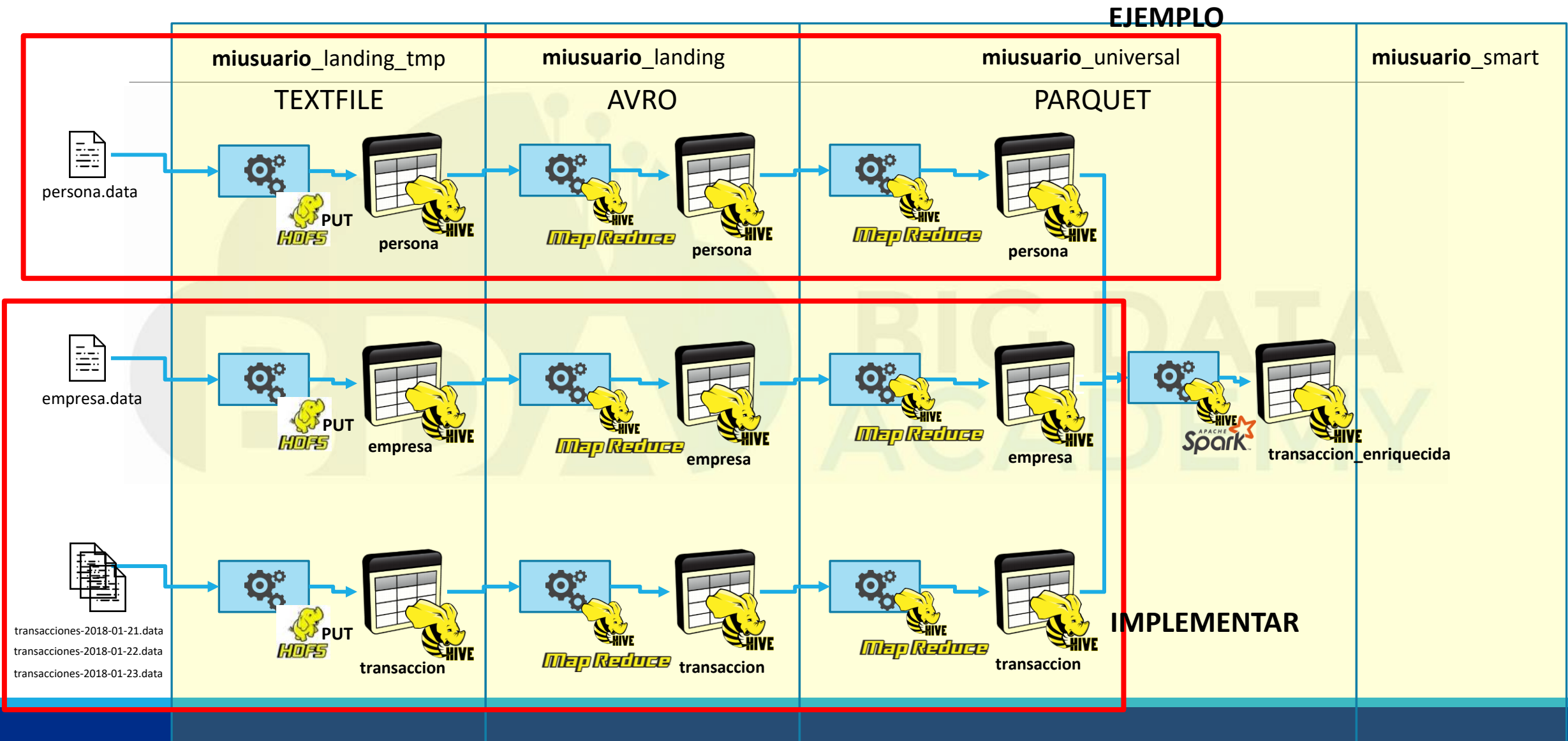


# Implementar un Datalake



# PASO 1: Crear la estructura de carpetas sobre HDFS

```
/user/miusuario
  /ejercicio2
    /database
      /miusuario_landing_tmp
        /persona
        /empresa
        /transaccion
      /miusuario_landing
        /persona
        /empresa
        /transaccion
      /miusuario_universal
        /persona
        /empresa
        /transaccion
        /transaccion_enriquecida
      /miusuario_smart
    /schema
      /database
        /miusuario_landing
```



ejemplo\_paso\_1.txt

# PASO 2: Crear las bases de datos

---



ejemplo\_paso\_2.txt

## PASO 3: Crear la capa LANDING\_TMP

---



ejemplo\_paso3.txt

# PASO 4: Crear la capa LANDING

---



ejemplo\_paso\_4.txt

# PASO 5: Crear la capa UNIVERSAL

---



ejemplo\_paso\_5.txt

---

# Ejercicio 2

---

CREAR UNA TABLA DESNORMALIZADA SOBRE UNIVERSAL



# Crear “transaccion\_enriquecida”

En la capa “universal” crear la tabla “transaccion\_enriquecida” la cual debe ser construida en función de las tablas “persona”, “empresa” y “transaccion”

Los campos de la tabla son los siguientes:

CAMPO	TIPO	DESCRIPCIÓN
ID_PERSONA	INT	ID de la persona que realizó la transacción
NOMBRE_PERSONA	STRING	Nombre de la persona que realizó la transacción
EDAD_PERSONA	INT	Edad de la persona que realizó la transacción
SALARIO_PERSONA	DOUBLE	Salario de la persona que realizó la transacción
TRABAJO_PERSONA	STRING	Nombre de la empresa donde trabaja la persona
MONTO_TRANSACCION	DOUBLE	Monto de la transacción realizada
FECHA_TRANSACCION	STRING	Fecha de la transacción realizada
EMPRESA_TRANSACCION	STRING	Nombre de la empresa donde se realizó la transacción



---

# Ejercicio 3

---

TUNING

BIG DATA  
ACADEMY

# Tuneo de un proceso HIVE

---

En la siguiente sesión...

