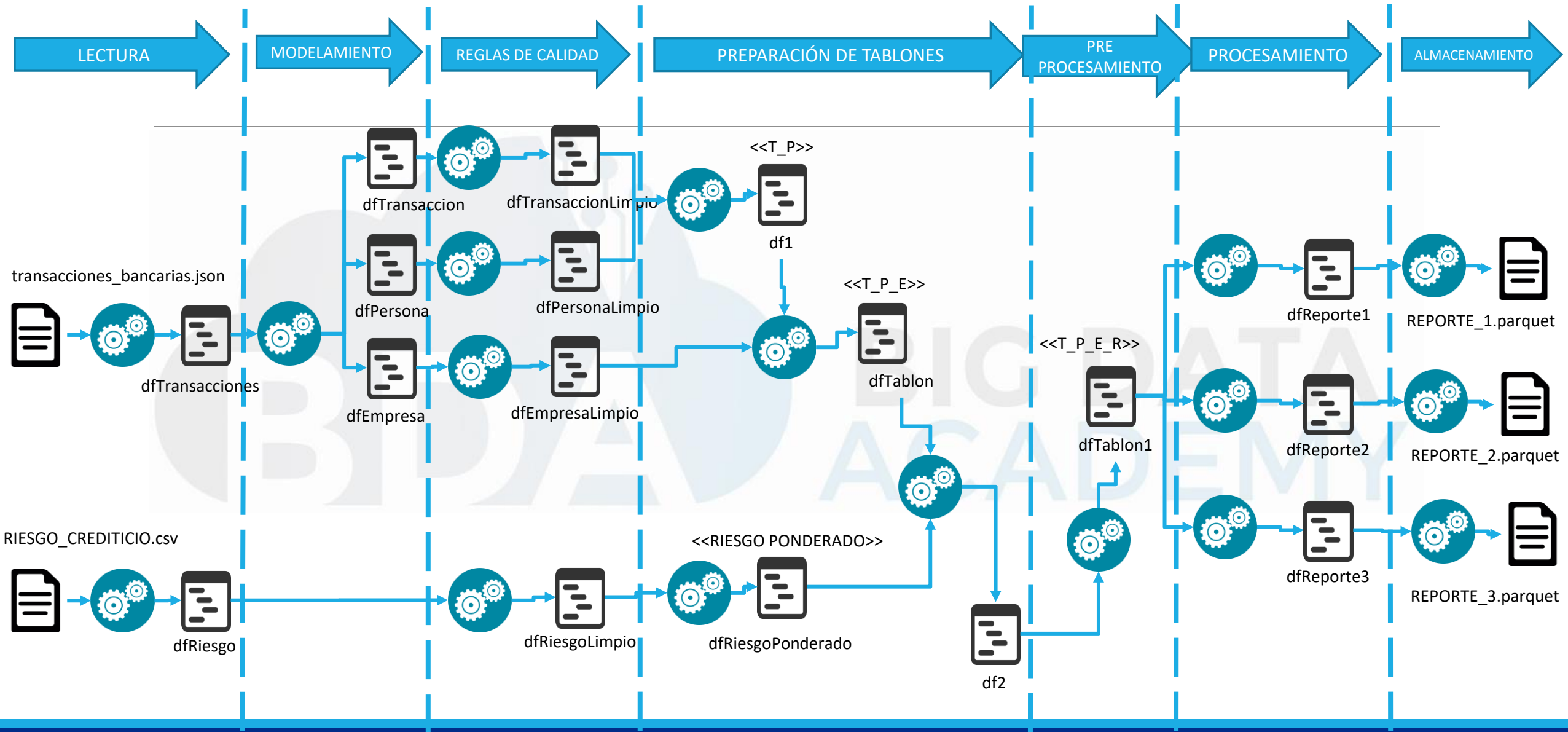

SPARK [EJERCICIO 4]

PROCESAMIENTO AVANZADO

EJERCICIO 4

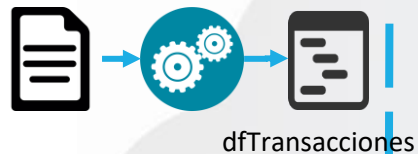
**IMPLEMENTAR EL
SIGUIENTE PROCESO**

Ejercicio 4



LECTURA

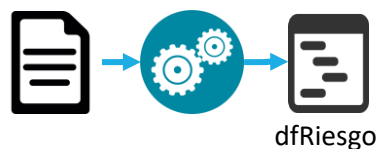
transacciones_bancarias.json



PASO 1: LECTURA

Leer los archivos de datos en
DATAFRAMES

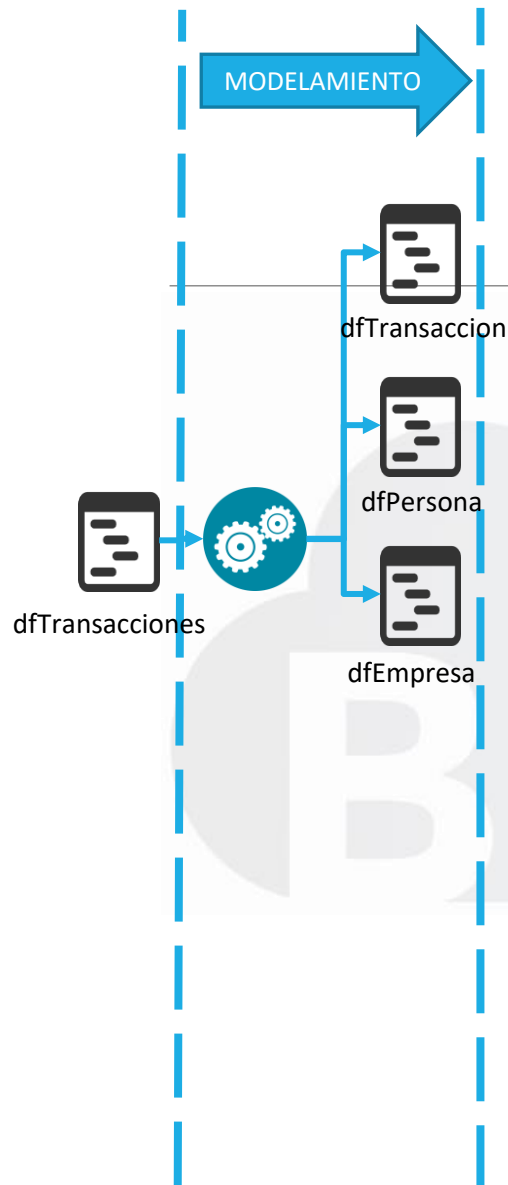
RIESGO_CREDITICIO.csv



MODELAMIENTO

PASO 2: MODELAMIENTO

Estructurar el dfTransacciones



dfTransaccion

ID_PERSONA
ID_EMPRESA
MONTO
FECHA

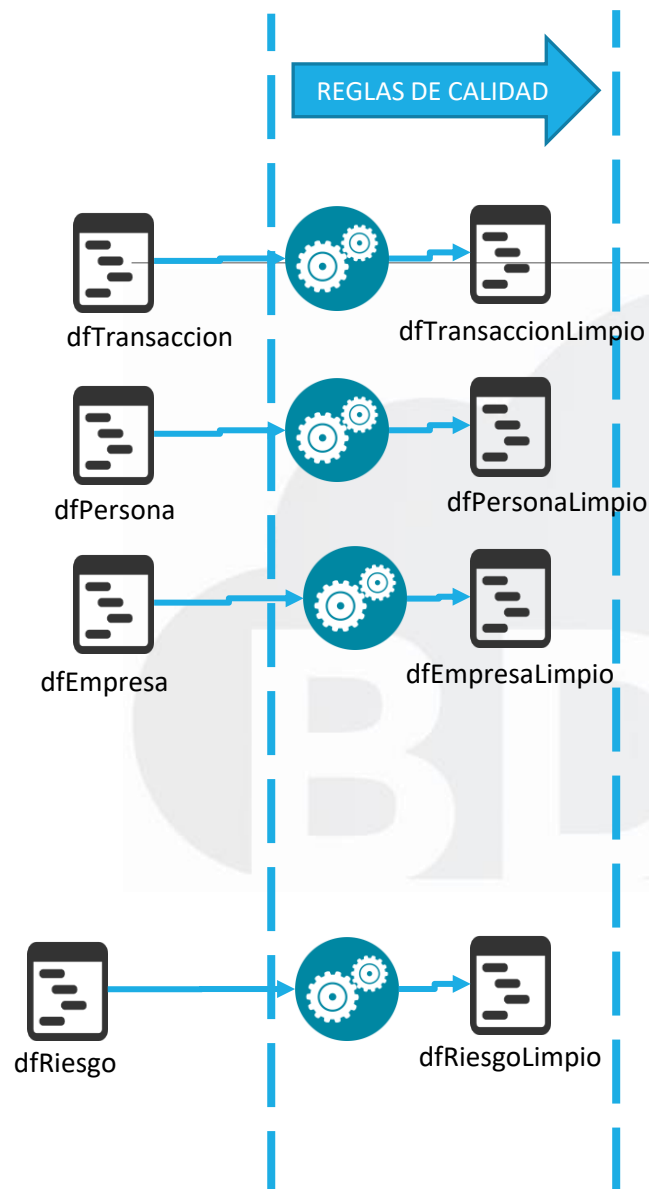
dfEmpresa

ID_EMPRESA
NOMBRE_EMPRESA

dfPersona

ID_PERSONA
NOMBRE_PERSONA
EDAD
SALARIO

PASO 3: REGLAS DE CALIDAD



PERSONA

ID_PERSONA	NO NULO
SALARIO	RANGO: [0, 100000>
EDAD	RANGO: <0, 60>

RIESGO

ID_CLIENTE	NO NULO
RIESGO_CENTRAL_1	RANGO: [0, 1]
RIESGO_CENTRAL_2	RANGO: [0, 1]
RIESGO_CENTRAL_3	RANGO: [0, 1]

EMPRESA

ID_EMPRESA	NO NULO
------------	---------

TRANSACCION

ID_PERSONA	NO NULO
ID_EMPRESA	NO NULO
MONTO	RANGO: [0, 100000>

PASO 4: CREACIÓN DE UDF



dfRiesgoPonderado
ID_CLIENTE
RIESGO_PONDERADO

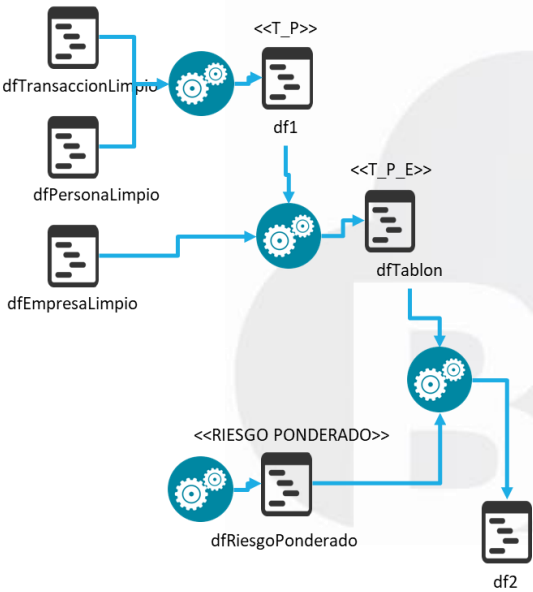
```
FUNCION calcularRiesgoPonderado(riesgo1, riesgo2, riesgo3):  
    resultado = 0
```

```
    resultado = (2 x riesgo1 + 3 x riesgo2 + 2 x riesgo3) / 7
```

```
    RETORNAR resultado
```

PASO 5: PREPARACIÓN DE TABLONES

PREPARACIÓN DE TABLONES



DATAFRAME: df1

CAMPO	PROVIENE DE
ID_PERSONA	dfTransaccionLimpio
NOMBRE_PERSONA	dfPersonaLimpio
EDAD_PERSONA	dfPersonaLimpio
SALARIO_PERSONA	dfPersonaLimpio
ID_EMPRESA_TRANSACCION	dfTransaccionLimpio
MONTO_TRANSACCION	dfTransaccionLimpio
FECHA_TRANSACCION	dfTransaccionLimpio

DATAFRAME: dfTablon

CAMPO	PROVIENE DE
ID_PERSONA	df1
NOMBRE_PERSONA	df1
EDAD_PERSONA	df1
SALARIO_PERSONA	df1
ID_EMPRESA_TRANSACCION	df1
NOMBRE_EMPRESA	dfEmpresaLimpio
MONTO_TRANSACCION	df1
FECHA_TRANSACCION	df1

DATAFRAME: df2

CAMPO	PROVIENE DE
ID_PERSONA	dfTablon
NOMBRE_PERSONA	dfTablon
EDAD_PERSONA	dfTablon
SALARIO_PERSONA	dfTablon
RIESGO_PONDERADO	dfRiesgoPonderado
ID_EMPRESA_TRANSACCION	dfTablon
NOMBRE_EMPRESA	dfTablon
MONTO_TRANSACCION	dfTablon
FECHA_TRANSACCION	dfTablon

PASO 6: PRE-PROCESAMIENTO

PRE PROCESAMIENTO

<<T_P_E>>



df2



dfTablon1

FILTRAMOS

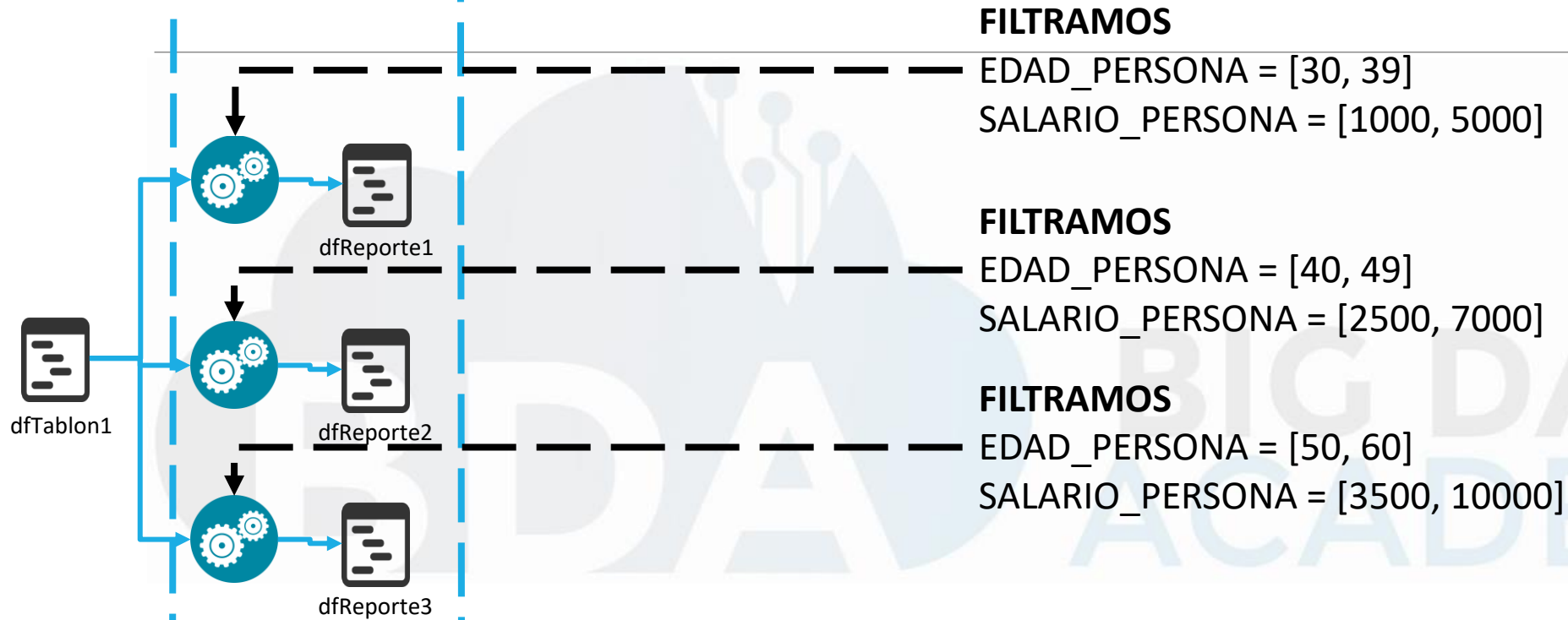
MONTO_TRANSACCION > 500

NOMBRE_EMPRESA == "Amazon"

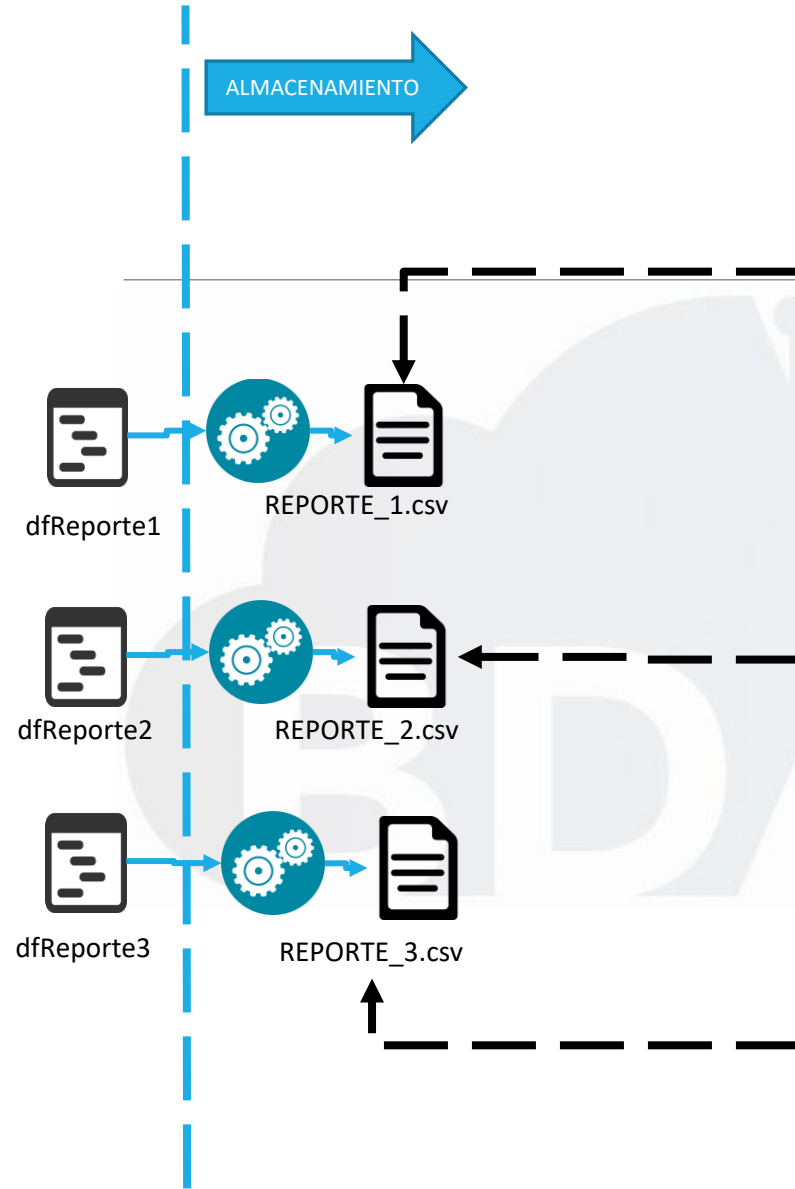
BIG DATA
ACADEMY

PROCESAMIENTO

PASO 7: PROCESAMIENTO



PASO 8: ALMACENAMIENTO



ALMACENAMIENTO

Formato: csv
 Cabecera: true
 Delimitador: |
 Encoding: ISO-8859-1
 Directorio: dbfs:///FileStore/_bigdata/output/EJERCICIO_3/REPORTE_1

ALMACENAMIENTO

Formato: csv
 Cabecera: true
 Delimitador: |
 Encoding: ISO-8859-1
 Directorio: dbfs:///FileStore/_bigdata/output/EJERCICIO_3/REPORTE_2

ALMACENAMIENTO

Formato: csv
 Cabecera: true
 Delimitador: |
 Encoding: ISO-8859-1
 Directorio: dbfs:///FileStore/_bigdata/output/EJERCICIO_3/REPORTE_3