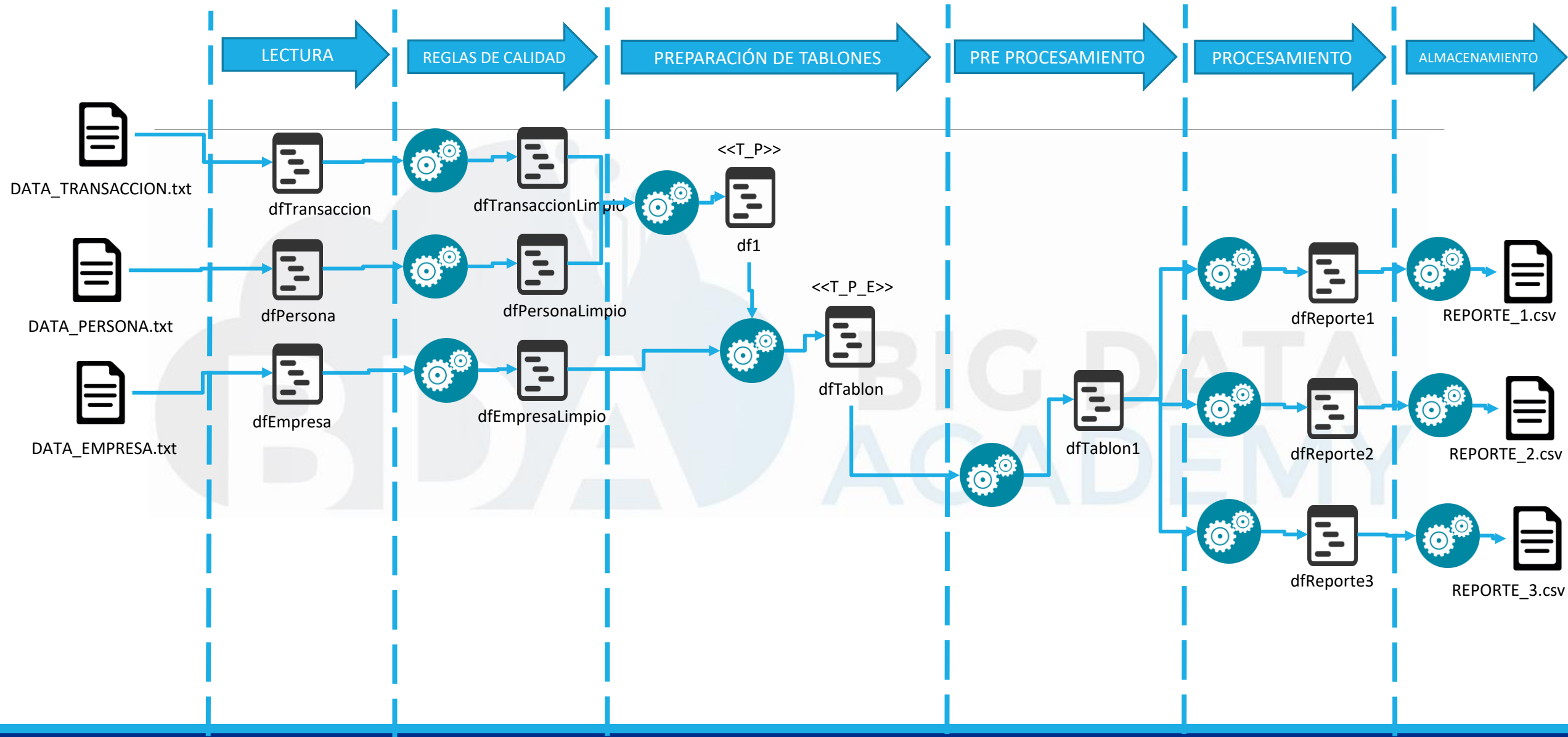

PYSPARK [EJERCICIO 3]

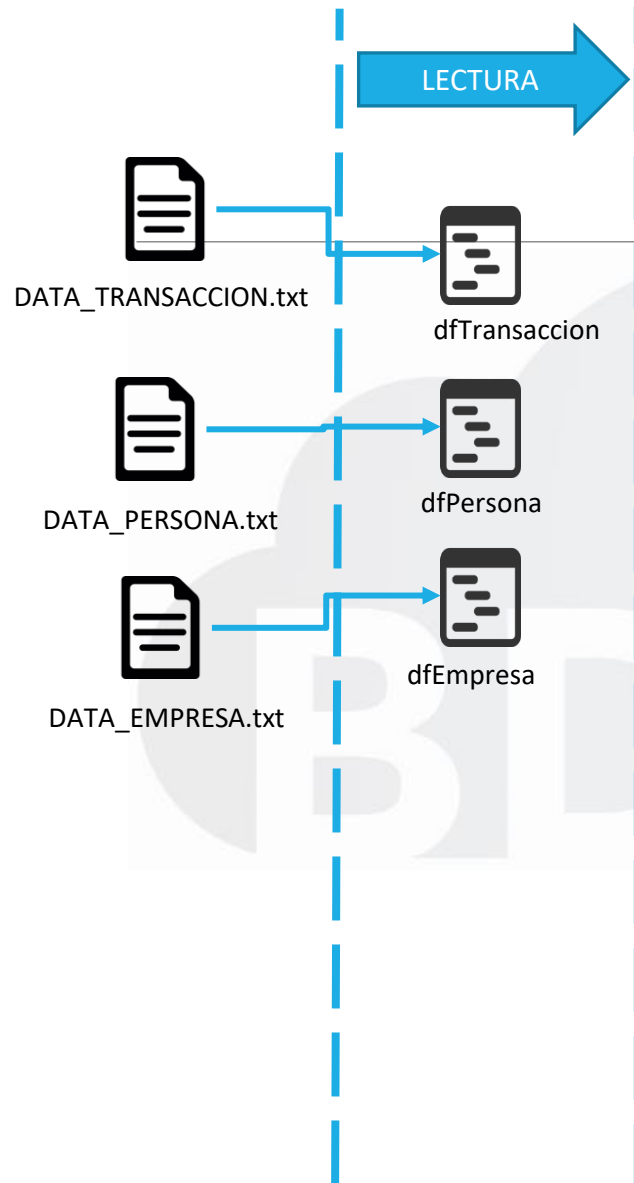
ARQUETIPO DE PROCESAMIENTO BÁSICO

EJERCICIO 3

**IMPLEMENTAR EL
SIGUIENTE PROCESO**

Ejercicio 3

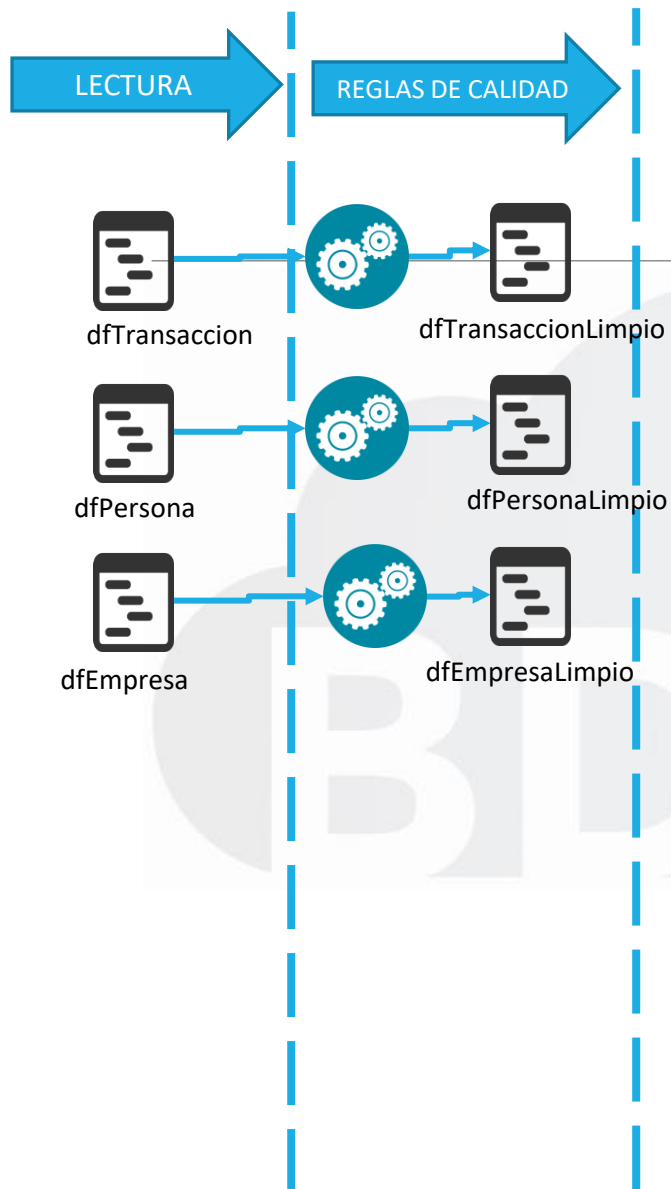




PASO 1: LECTURA

Leer los archivos de datos en
DATAFRAMES

PASO 2: REGLAS DE CALIDAD



PERSONA

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

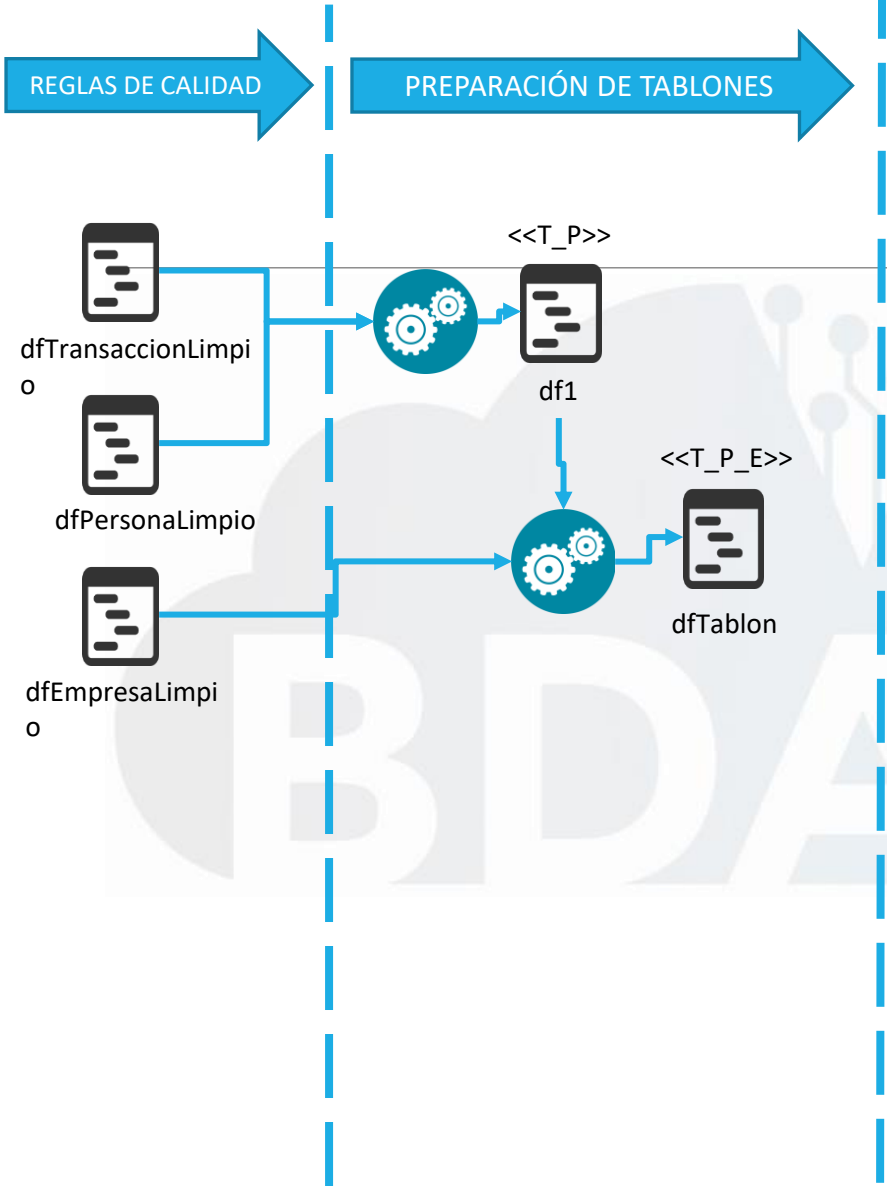
EMPRESA

ID	NO NULO
----	---------

TRANSACCION

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

PASO 3: PREPARACIÓN DE TABLONES



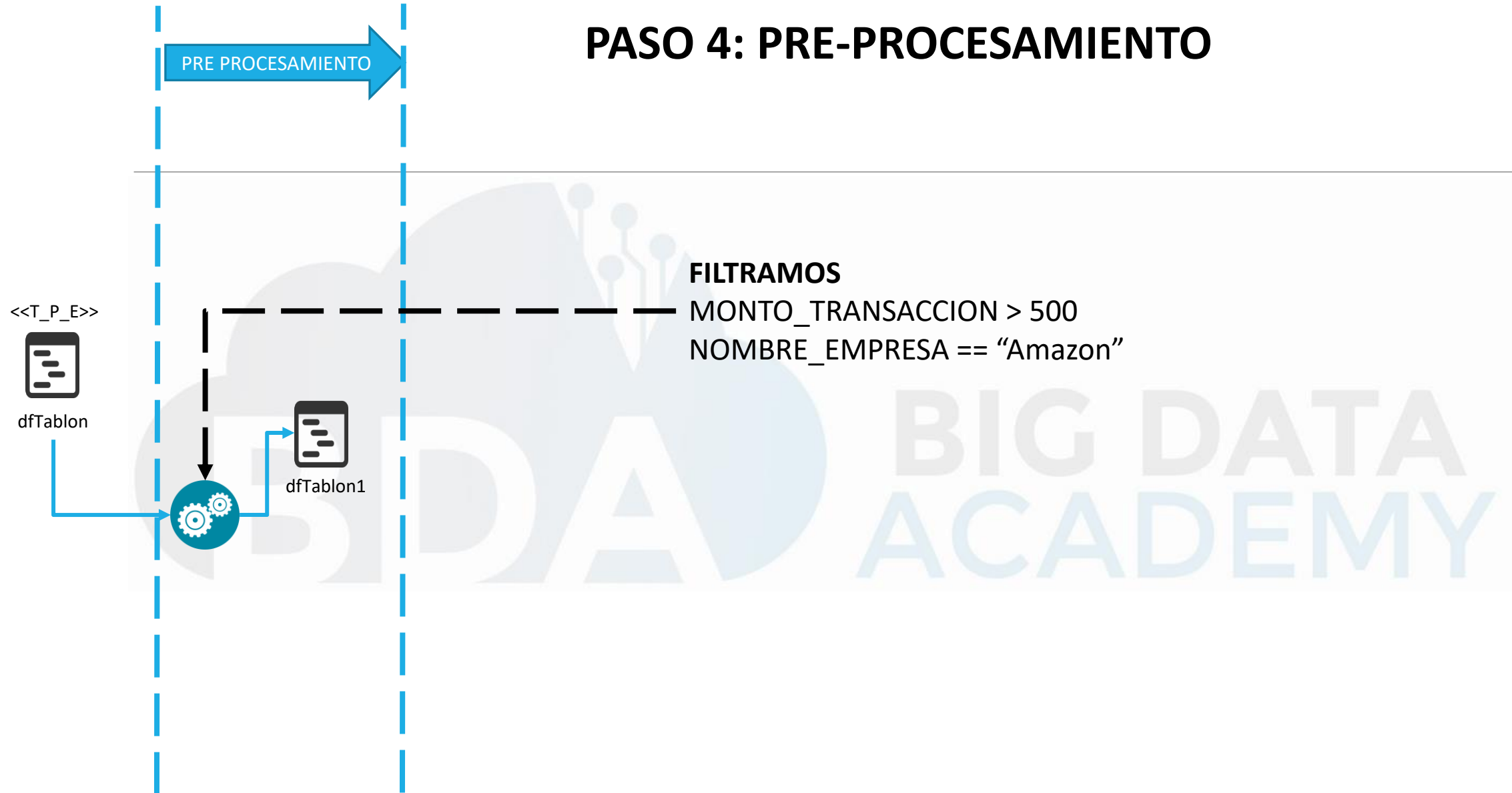
DATAFRAME: df1

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

DATAFRAME: dfTablon

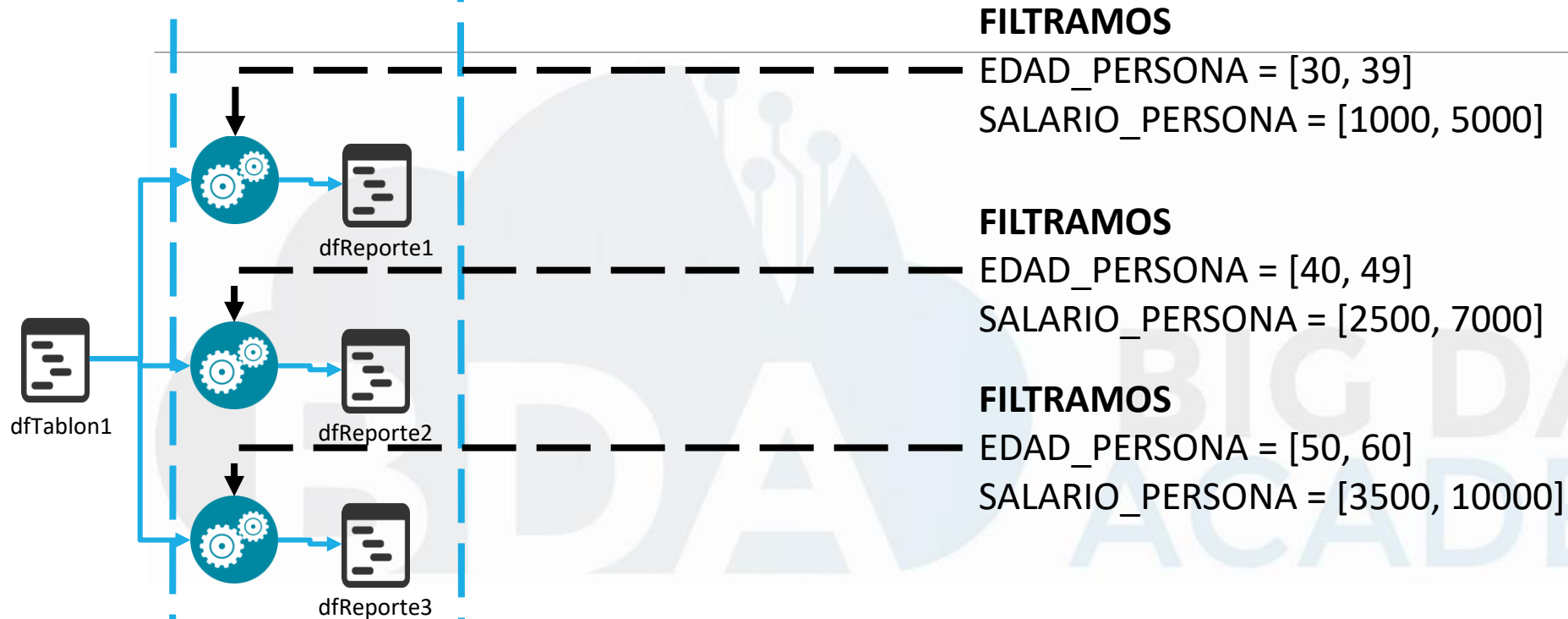
CAMPO	PROVIENE DE
ID_PERSONA	df1
NOMBRE_PERSONA	df1
EDAD_PERSONA	df1
SALARIO_PERSONA	df1
ID_EMPRESA_PERSONA	df1
ID_EMPRESA_TRANSACCION	df1
NOMBRE_EMPRESA	dfEmpresaLimpio
MONTO_TRANSACCION	df1
FECHA_TRANSACCION	df1

PASO 4: PRE-PROCESAMIENTO

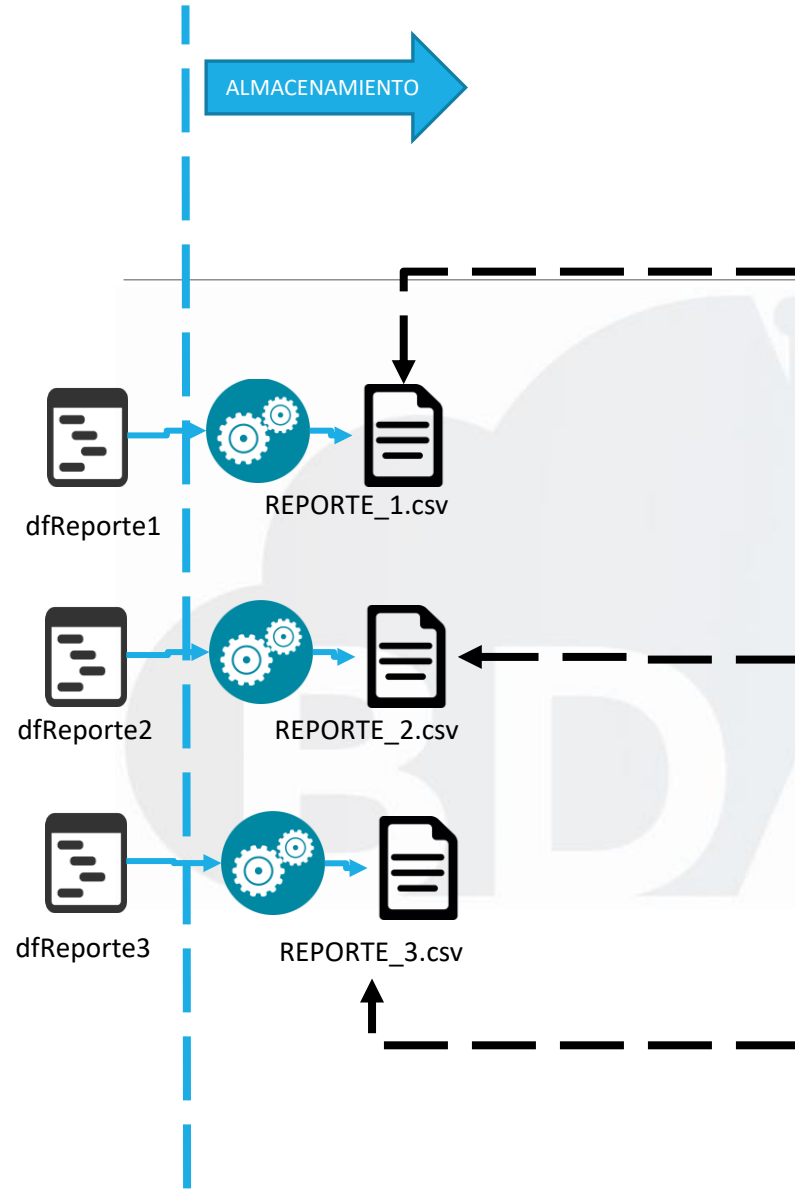


PROCESAMIENTO

PASO 5: PROCESAMIENTO



PASO 6: ALMACENAMIENTO



ALMACENAMIENTO

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

ALMACENAMIENTO

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

ALMACENAMIENTO

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