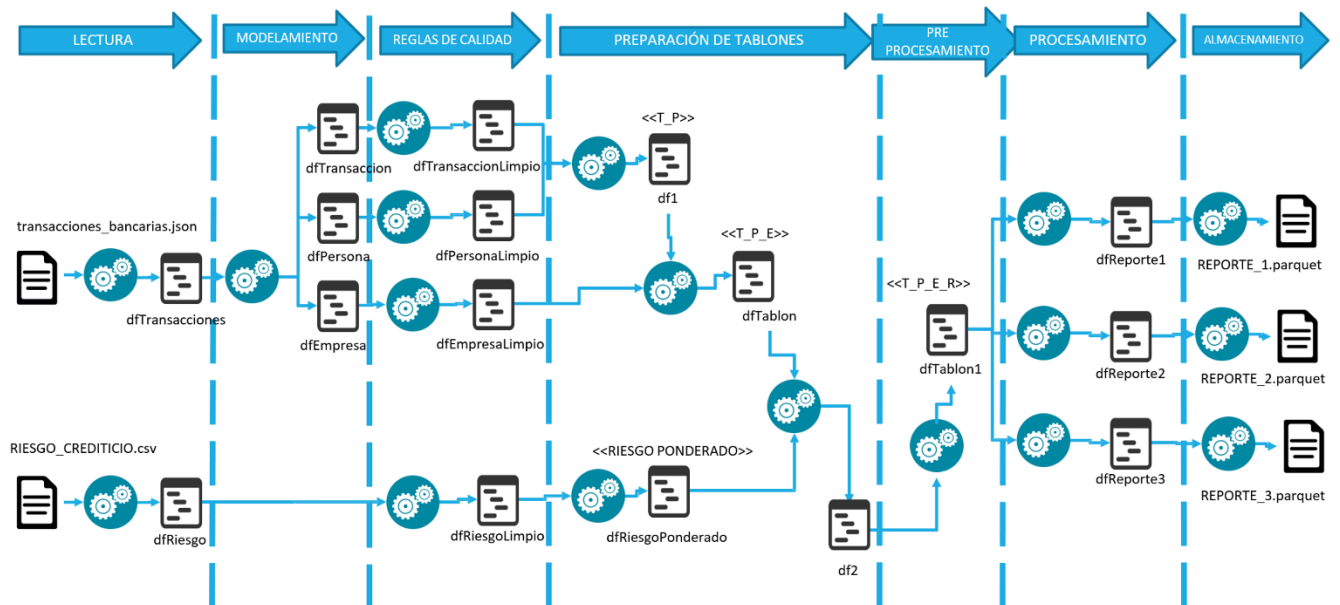


## Hands On Integrador - Arquetipo de Procesamiento Avanzado

Implementar el siguiente proceso lógico en Spark

**Importante:** Reservar el 10% de la potencia del siguiente clúster:  
1000GB RAM | 400 uCPU

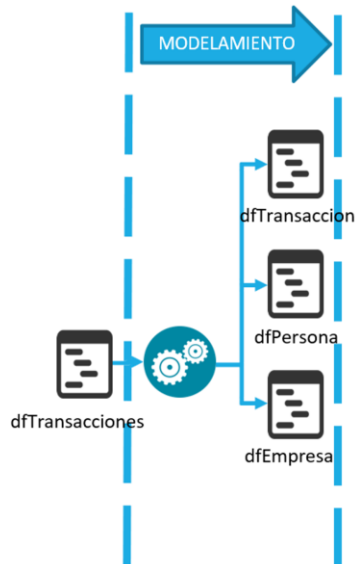


## Explicación – Paso a Paso



### Paso 1: Lectura

Leer los archivos de datos en  
DATAFRAMES



## Paso 2: Modelamiento

Estructurar el dfTransacciones

### dfTransaccion

ID_PERSONA
ID_EMPRESA
MONTO
FECHA

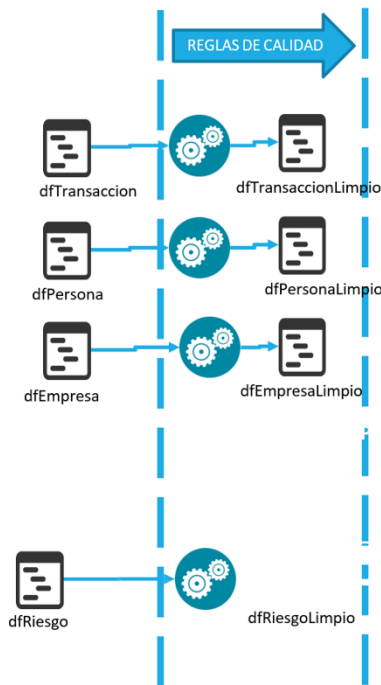
### dfPersona

ID_PERSONA
NOMBRE_PERSONA
EDAD
SALARIO

ID_EMPRESA
------------

NOMBRE_EMPRESA
----------------

## 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	

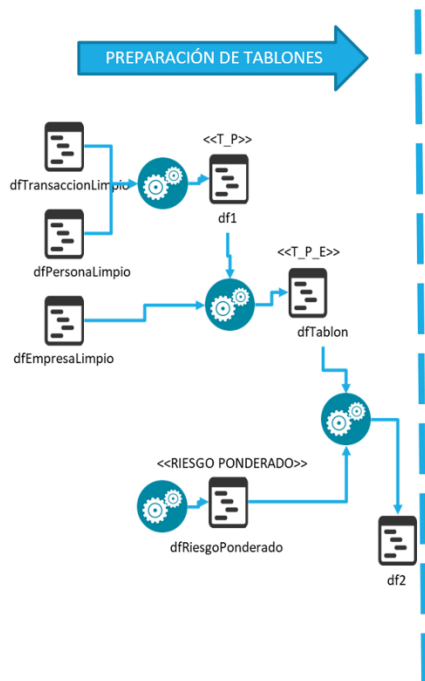
FUNCION **calcularRiesgoPonderado**(riesgo1, riesgo2, riesgo3):

resultado = 0

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

RETORNAR resultado

# Paso 5: Preparación de Tablones



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

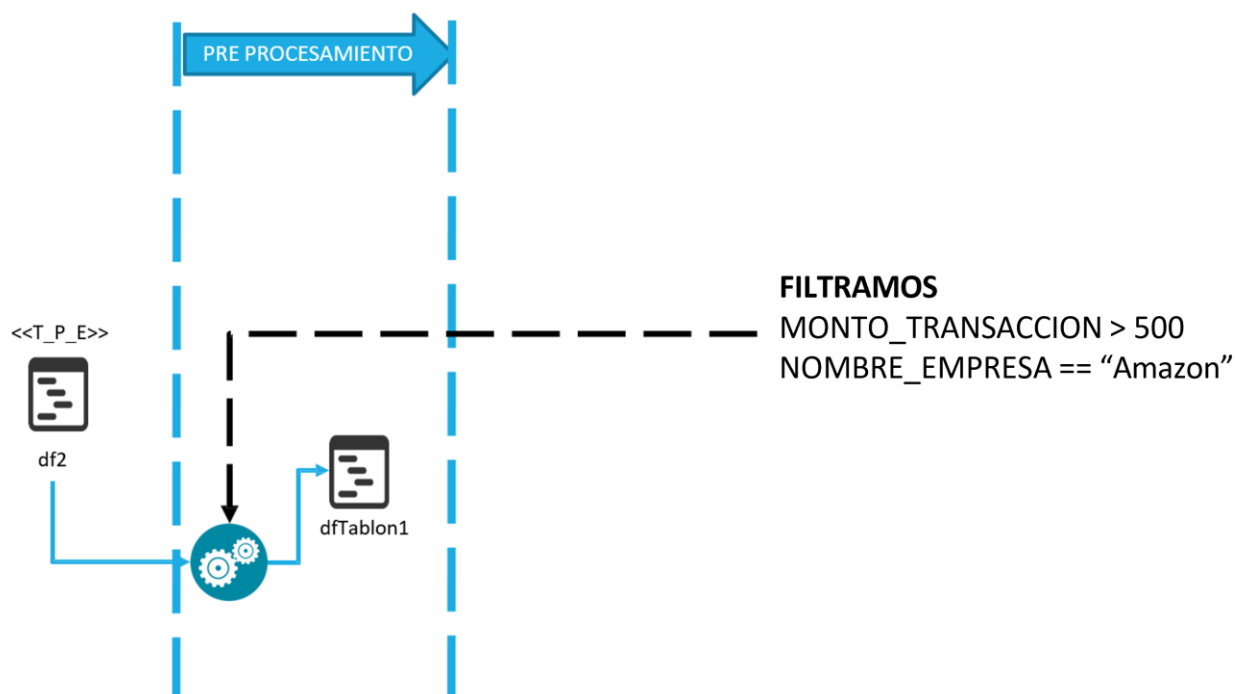
DATAFRAME: dfTablon

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

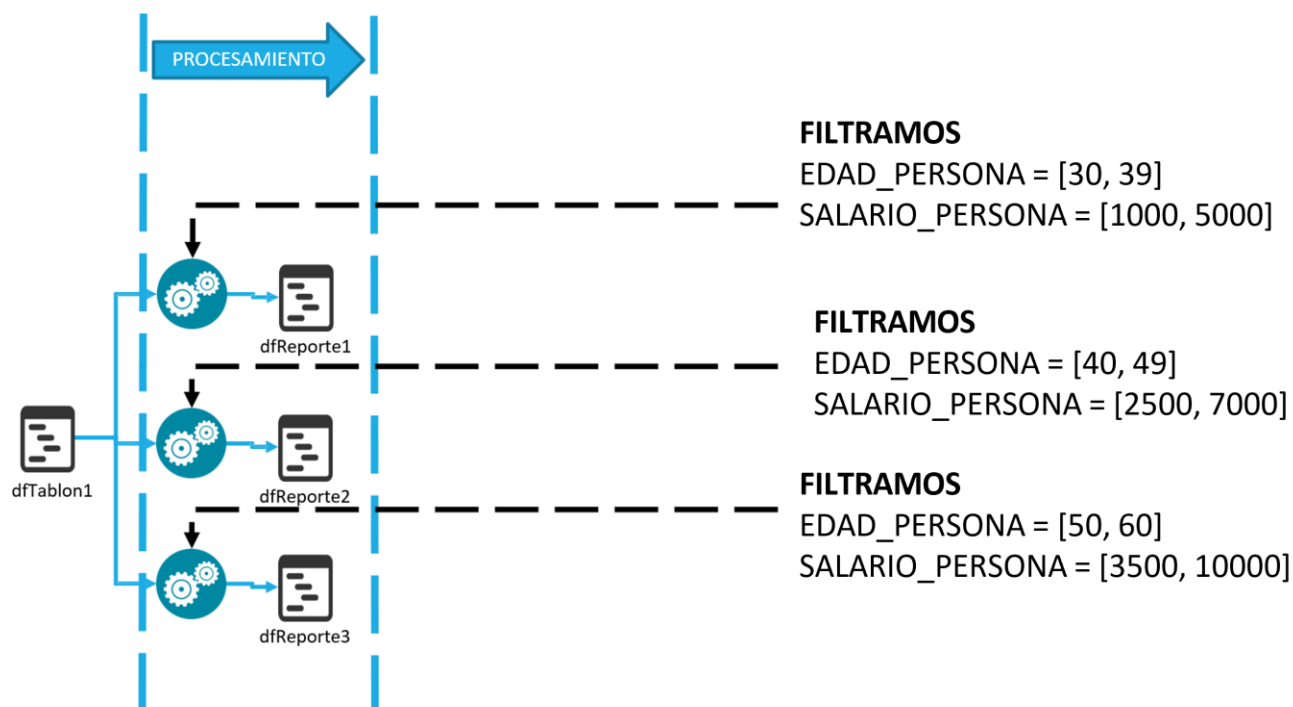
DATAFRAME: df2

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



## Paso 7: Procesamiento



## Paso 8: Almacenamiento

