# Automatización del proceso SWIFT Parsing y reprocesamiento de históricos

## Propuesta de integración con Control-M

El presente documento describe el proceso actual de ingesta y reprocesamiento de pagos SWIFT, así como una propuesta para su automatización mediante Control-M, con el objetivo de facilitar la ejecución flexible y escalable para nuevos BICs y reprocesamientos históricos.

## 1. Contexto actual

El proceso SWIFT Parsing se encarga de leer los ficheros descargados manualmente desde SWIFT Alliance Warehouse (ZIP + CSV), parsear los mensajes y generar ficheros Parquet que convergen en un esquema común y se almacenan en un HDFS. Este sistema de ficheros distribuido está particionado por fecha y existen carpetas mensuales (YYYYMM), cada una contiene los parquets generados por el proceso respectivos al mes de procesamiento correspondiente.

## 2. Estrategias de reprocesamiento para el BIC de Argentina

| Opción | Descripción | Ventajas | Inconvenientes |
| --- | --- | --- | --- |
| A. Ejecuciones manuales mensuales/quincenales | Descargar de AW el ZIP/CSV del periodo y ejecutarlo manualmente con SWIFT\_parsing.py. | Trazabilidad perfecta, coherencia con el histórico. | Requiere múltiples ejecuciones (24-48 para 2 años). |
| B. Ejecución única y desgranado posterior | Generar un único Parquet global de 2 años y luego repartir transacciones por mes en HDFS. | Solo una ejecución, más rápida. | Rompe trazabilidad y requiere código adicional. |

La opción A es la recomendada actualmente, añadiendo el sufijo \_3 para reprocesamientos o nuevos BICs (como Argentina).

**3. Propuesta de automatización mediante la malla de Control-M**

El objetivo es ejecutar el proceso SWIFT Parsing de forma parametrizable desde Control-M, permitiendo seleccionar BICs, fechas, tipo de ejecución, granularidad temporal y otros parámetros sin intervención manual.

### Parámetros de ejecución recomendados

| **Parámetro** | **Descripción funcional** | **Ejemplo / Formato esperado** | **Uso dentro del proceso** |
| --- | --- | --- | --- |
| **EXEC\_MODE** | Tipo de ejecución: NORMAL (ingesta mensual/quincenal) o REPROCESS (histórico). | NORMAL / REPROCESS | Permite distinguir en logs y rutas si se trata de una reingesta histórica. |
| **BIC\_LIST** | Lista de BICs a procesar (propios o de interés). | "BICARABCXXX,BICUSXXX,..." | Filtrado o identificación posterior (se puede usar para seleccionar qué ficheros procesar). |
| **START\_DATE** | Fecha inicial del rango de reprocesamiento. | 2024-01-01 | Determina desde qué mes/quincena se inicia el reprocesamiento. |
| **END\_DATE** | Fecha final del rango de reprocesamiento. | 2025-12-31 | Determina hasta qué mes/quincena se ejecuta. |
| **PERIODICITY** | Nivel de granularidad de los ficheros: MONTHLY o BIWEEKLY. | MONTHLY | Define si los ficheros tienen sufijo \_1/\_2 o \_3. |
| **TARGET\_ENV** | Entorno destino: PRE, PRO, TEST. | PRO | Define las rutas HDFS o Kerberos a utilizar. |
| **DAT\_DIR** | Ruta local temporal donde Control-M dejará el ZIP y el CSV a procesar. | /data/swift/input/ | Usado por el script SWIFT\_parsing. |
| **LOG\_LEVEL** | Nivel de log del proceso (INFO, DEBUG, ERROR). | INFO | Control de verbosidad. |
| **MERGE\_POLICY** | Indica si el parquet mensual debe sobrescribirse o concatenarse. | OVERWRITE / APPEND | Define cómo actualizar los merges en HDFS. |
| **FILE\_SUFFIX** | Sufijo numérico (\_1, \_2, \_3, etc.). | 3 | Marca la naturaleza del ZIP (quincenal o mensual extra). |

Cada ejecución de Control-M podrá configurar los parámetros según el caso de uso, permitiendo reprocesar históricos, añadir nuevos BICs o lanzar ingestas mensuales automáticas.

## 4. Conclusión

El proceso manual actual (Opción A) es robusto y alineado con la estructura histórica del proyecto, ya que mantiene la relación directa entre los ficheros brutos (ZIP/CSV) y los parquets finales generados en HDFS por mes.

Para el caso de Argentina, esta metodología permite incorporar fácilmente los nuevos ficheros (\_3) sin afectar los históricos \_1 y \_2, garantizando que:

* Cada mes conserva los consumos originales en bruto y dos parquet mensuales consolidados (Payments y Trades).
* El merge final se genera de forma controlada, sin sobrescribir los datos previos.
* Se mantiene la trazabilidad completa desde la fuente SWIFT Alliance Warehouse hasta los datos finales en HDFS.

En cambio, la Opción B (ejecución única y desgranado posterior) rompería esta trazabilidad, al mezclar en una sola ejecución datos de distintos meses y crear una dependencia artificial entre el parquet global y las carpetas mensuales. Esto dificultaría el proceso.

La parametrización propuesta permitirá flexibilidad, trazabilidad y escalabilidad en futuros reprocesamientos o incorporaciones de nuevos BICs.