

Interpretación de Resultados del Comparador

Propósito. Este documento explica, cómo interpretar las salidas del comparador entre una **tabla de referencia (REF)** y una **tabla nueva (NEW)**. Incluye definiciones de métricas, tipos de duplicados, opciones de configuración.

Incluyo en esta explicación dos casos reales de pruebas de ejecuciones en PRO desde el 2025-09-18 hasta 2025-09-24. La comparación es de REF contra REF. Se utilizó la siguiente **clave compuesta**:

geo, uetr, type, time, in_out, segregation, sender_reference

1) Tablas de salida y para qué sirven

1. Tabla Ejecutiva: bd.results_summary

A ^B _C block	A ^B _C metric	A ^B _C universe	A ^B _C numerator	A ^B _C denominator	A ^B _C pct
KPIS	Unique IDs	REF	55058	-	-
KPIS	Unique IDs	NEW	55058	-	-
KPIS	Total rows REF	ROWS	55841	-	-
KPIS	Total rows NEW	ROWS	55841	-	-
KPIS	Total diff(new-ref)	ROWS	0	55841	0.0000%
KPIS	Global quality	REF	55050	55058	99.9855%
EXACT MATCH	1:1 (all columns)	BOTH	55058	55058	100.0000%
PARTIAL MATCH	1:1 (match & no_match co...	BOTH	0	55058	0.0000%
GAP	1:0 (only in ref)	REF	0	55058	0.0000%
GAP	0:1 (only in new)	NEW	0	55058	0.0000%
DUPS	duplicates (both)	BOTH	8	55058	0.0145%
DUPS	duplicates (only in ref)	REF	0	55058	0.0000%
DUPS	duplicates (only in new)	NEW	0	55058	0.0000%

Resumen ejecutivo de la ejecución. No muestra detalle por fila. Incluye:

BLOQUE 1:

- **Unique IDS:** número de identificadores únicos según la clave definida.
- **Total rows:** conteo de filas totales que llegan (incluye duplicados).
- **Total diff:** diferencia de filas entre NEW y REF ($NEW - REF$).
- **Global Quality:** métrica global de calidad. **Si existen duplicados no alcanza el 100%**, porque los dupes forman parte de su cálculo.
- **Samples por métrica (hasta 6 IDs):** para cada métrica relevante (p. ej., diferencias y duplicados) se guardan **hasta 6 identificadores de ejemplo**. Sirven para localizar rápidamente esos casos tanto en `results_differences` como en `results_duplicates`.

BLOQUE 2:

Tabla de Diferencias: `bd.results_differences`: `bd.results_differences**`

Detalle de las divergencias a nivel de identificador y columnas. Puede incluir (según configuración): - **EXACT_MATCH (1:1):** todas las columnas coinciden (se muestra si está activada la opción de guardar exactos).

- **PARTIAL_MATCH (1:1):** el identificador coincide, pero hay diferencias en una o más columnas.

- En esta situación, pueden guardarse **MATCH** y **NO_MATCH** si la opción está habilitada. -

GAP: elementos presentes en una tabla y ausentes en la otra. Se muestran **hasta 6 ejemplos** de identificadores por tipo de hueco.

BLOQUE 3:

Tabla de Duplicados: `bd.results_duplicates`

Detalle de duplicidades de identificadores:

- **both:** el mismo identificador se duplica **en REF y en NEW**. *Nota: cuando ocurre en ambos orígenes, se descuenta de los contadores independientes de cada lado.*

- **duplica REF:** duplicado solo en la referencia.

- **duplica NEW:** duplicado solo en la nueva.

Además, se distinguen dos subtipos en la tabla de duplicados:

- **Duplicados exactos:** filas idénticas.

- **Duplicados con variación:** el identificador coincide, pero alguna(s) columna(s) difiere(n). Es habitual cuando algún campo usado como identificador viene **NULL**.

La columna **variaciones** muestra el detalle de la diferencia por columna, p. ej.:

`bic_1: [valor_1, valor_2]` (y así sucesivamente para el resto de los campos afectados).

Nota de producto: En el futuro, es posible unificar la categoría visual de duplicados en una sola etiqueta (**both**), coherente con el resumen.

2) Existen 2 Casos de métricas que se dan en PRO (casos reales)

Caso 1 — Igualdad perfecta (2025-09-21)

- **Summary** igualdad total. **KPI Global quality** y **EXACT MATCH** estarán al **100%**
- El resto de las métricas deben estar a 0.

A ^B C block	A ^B C metric	A ^B C universe	A ^B C numerator	A ^B C denominator	A ^B C pct	A ^B C sample
KPIS	Unique IDs	REF	1572	-	-	-
KPIS	Unique IDs	NEW	1572	-	-	-
KPIS	Total rows REF	ROWS	1572	-	-	-
KPIS	Total rows NEW	ROWS	1572	-	-	-
KPIS	Total diff(new-ref)	ROWS	0	1572	0.0000%	-
KPIS	Global quality	REF	1572	1572	100.0000%	-
EXACT MATCH	1:1 (all columns)	BOTH	1572	1572	100.0000%	> BR_000
PARTIAL MATCH	1:1 (match & no_match co...	BOTH	0	1572	0.0000%	-
GAP	1:0 (only in ref)	REF	0	1572	0.0000%	-
GAP	0:1 (only in new)	NEW	0	1572	0.0000%	-
DUPS	duplicates (both)	BOTH	0	1572	0.0000%	-
DUPS	duplicates (only in ref)	REF	0	1572	0.0000%	-
DUPS	duplicates (only in new)	NEW	0	1572	0.0000%	-

- **Differences:** solo veríamos **EXACT_MATCH** si la opción de guardarlos estuviera activada; si no, la tabla queda vacía.

Table	A ^B C column	A ^B C value_ref	A ^B C value_n...	A ^B C results
1626492-470b-46c8-afa7-552194f657d6_103_09:29_I_BR_MBB17D386535C0A	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
17a2331-2564-4261-bceb-f89b5d157be2_103_10:49_I_BR_MBB4F727950CE...	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
8fc9b6-c2cc-41bc-92e2-35d776c3106f_103_15:04_I_BR_MBB81ECE45A2181	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
9989e4-1a76-4b04-90c6-79582a38016a_103_08:46_I_ES_1560696010018553	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
590410-919b-4933-8b78-96a88af7fbbd_103_16:06_O_ES_2025092200079985	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
K_ae7149ef-7fce-4859-ae91-95353962aed5_103_18:45_O_MX_49986960100...	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
db065c-263a-493b-8d80-bbdb0dd16fee_103_05:08_I_PL_2509210856947686	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
0fd9e0-a14f-4922-ae72-fce50be83ae1_103_23:09_I_PL_2509210868376233	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
0f11eb-58ba-4b56-a6f8-56880137d106_103_18:48_I_ES_2354696010056626	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
0d9313-f95a-488c-a50e-ff6c96c9644b_103_16:03_I_ES_0950696010011159	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
1fdffb-ea00-4a22-b2e9-5ffb3199fa5e_103_12:11_I_ES_1736696010045921	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
d6d32c-cb56-4eec-927c-3a4eff766241_103_03:49_O_ES_MBB2B1A27CB9CC9	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
1f8749c-4c5e-4144-bdb0-5c0e6c45c12a_103_12:18_O_MX_D0452650026901	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
9ff9147-98a6-4169-af1b-eafb35b6d782_103_10:46_I_MX_OPAY25092100027	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
c774a2e-1bab-4460-bea2-c41c01f4af0f_103_22:55_I_MX_OPAY25092000141	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH

Duplicates: debe estar vacía.

Caso 2 — Escenario típico (funcionalmente correcto) (2025-09-24)

- **Global Quality ≈ 99.9%:** no llega al 100% porque hay duplicados.
- Se observan **8 IDs únicos** con ejemplos duplicados en **ambos orígenes** (ver detalle y tipología en results_duplicates).

$A_C^{B_C}$ block	$A_C^{B_C}$ metric	$A_C^{B_C}$ universe	$A_C^{B_C}$ numerator	$A_C^{B_C}$ denominator	$A_C^{B_C}$ pct	$A_C^{B_C}$ sample
KPIS	Unique IDs	REF	55058	-	-	-
KPIS	Unique IDs	NEW	55058	-	-	-
KPIS	Total rows REF	ROWS	55841	-	-	-
KPIS	Total rows NEW	ROWS	55841	-	-	-
KPIS	Total diff(new-ref)	ROWS	0	55841	0.0000%	-
KPIS	Global quality	REF	55050	55058	99.9855%	-
EXACT MATCH	1:1 (all columns)	BOTH	55058	55058	100.0000%	> BJC_23
PARTIAL MATCH	1:1 (match & no_match co...	BOTH	0	55058	0.0000%	-
GAP	1:0 (only in ref)	REF	0	55058	0.0000%	-
GAP	0:1 (only in new)	NEW	0	55058	0.0000%	-
DUPS	duplicates (both)	BOTH	8	55058	0.0145%	> FRK_N
DUPS	duplicates (only in ref)	REF	0	55058	0.0000%	-
DUPS	duplicates (only in new)	NEW	0	55058	0.0000%	-

- **Differences:** muestra PARTIAL_MATCH (si procede) y, si la opción está activa, también EXACT_MATCH.
- **Duplicados:** muestra los duplicados por origen (ref | new) y el identificador que hayamos decidido.

$A_C^{B_C}$ origin	$A_C^{B_C}$ id	$A_C^{B_C}$ exact_duplicates	$A_C^{B_C}$ dupes_w_variatio...	$A_C^{B_C}$ occurrences	$A_C^{B_C}$ variations
new	MUN_NOT_PROVIDED_103_0812_UJT_NULL	0	170	171	> bic_1: [BPPITRR,CIPBITMM,SUNITMM] sender_bic: [BPPITRR,CIPBITMM,SUNITMM] session_sequence: [02b8085-7527-4a85-a3da-a375e117c6b,034ed58-6f76-4395-a0e9-...
new	MUN_NOT_PROVIDED_103_0812_UJT_NULL	0	578	579	> bic_1: [BACIT22,BAPIT22,BITATRR,BPMOT22,BPPITRR,CIPBITMM,CACIDEDDOXX,CRPIT2P,CRAIT2P,GEBAEBB,ICRAITRR,PASCITMM,RVUALTZV,SELBIT22,UBSWCHZ,UNCL...
new	MUN_RCT_NULL_0614_UJT_SCL	0	1	2	> bic_1: [BPPITRR,CIPBITMM,SUNITMM] sender_bic: [BPPITRR,CIPBITMM,SUNITMM] session_sequence: [00000054958/20250924061405850,00000054966/20250924061405772] priority: [BPPITRRDOXX,CRPIT2P] transaction_instructed_amount: [5.750000,5.960000] transaction_instructed_amount_eur: [5.750000,5.960000] ordering_customer_name: [5.75.5.96] ordering_institution_bic: [CALAMANDRE] VALTER,TUMABARELLI MARTIN NAELI,INA ordering_institution_name: [BPPITRRDOXX,CRPIT2P] source: [DEFA_BITC03D960A03F4162BADEA09625D0,SEPA_31BB3A34622346FB3EBC33CF63C4B1]
new	FRK_NOTPROVIDED_103_2012_L_DE_NULL	0	1	2	> bic_1: [AARBDESXXXX,FMBKDEMMXXX] sender_bic: [AARBDESXXXX,FMBKDEMMXXX] session_sequence: [bc08b81-b3be-48df-a28c-079034ee198,9202467f-ce7a-457e-ba7...
new	MUN_NOTPROVIDED_103_1408_UJT_NULL	0	10	11	> bic_1: [BBAESMM,CRAIT2R,DEUTDEF,PASCITMM,RVUALTZV] sender_bic: [BBAESMM,CRAIT2R,DEUTDEF,PASCITMM,RVUALTZV] session_sequence: [36e7aa2-16a8-4582-...
new	> PT_85f7062-235d-41df-8367-87ad17d6...	0	1	2	> session_sequence: [9944784720,9944784722] bic_2: [CGDPTLXXXX,CHASUS330000] receiver_bic: [CGDPTLXXXX,CHASUS330000] mur: [TOTA509241604240,TOTA5092416...
new	MUN_NOTPROVIDED_103_1508_UJT_NULL	0	16	17	> bic_1: [BBRUBEC,ICRAITRR,SUNITMM,RVUALTZV] sender_bic: [BBRUBEC,ICRAITRR,SUNITMM,RVUALTZV] session_sequence: [00de965d-9dda-4412-a269-42b97a834b8,0cf...
new	MUN_NOTPROVIDED_103_1011_UJT_NULL	0	6	7	> bic_1: [DEUTDEF,PASCITMM,RZSBIT28] sender_bic: [DEUTDEF,PASCITMM,RZSBIT28] session_sequence: [2a3a91fc-2074-477d-8798-efbcb3a6206,4071f1ea-d0b1-4389-8440-...
ref	> PT_85f7062-235d-41df-8367-87ad17d6...	0	1	2	> session_sequence: [9944784720,9944784722] bic_2: [CGDPTLXXXX,CHASUS330000] receiver_bic: [CGDPTLXXXX,CHASUS330000] mur: [TOTA509241604240,TOTA5092416...
ref	MUN_NOTPROVIDED_103_1011_UJT_NULL	0	6	7	> bic_1: [DEUTDEF,PASCITMM,RZSBIT28] sender_bic: [DEUTDEF,PASCITMM,RZSBIT28] session_sequence: [2a3a91fc-2074-477d-8798-efbcb3a6206,4071f1ea-d0b1-4389-8440-...
ref	MUN_RCT_NULL_0614_UJT_SCL	0	1	2	> bic_1: [BPPITRR,CIPBITMM,SUNITMM] sender_bic: [BPPITRR,CIPBITMM,SUNITMM] session_sequence: [00000054958/20250924061405850,00000054966/20250924061405772] priority: [BPPITRRDOXX,CRPIT2P] transaction_instructed_amount: [5.750000,5.960000] transaction_instructed_amount_eur: [5.750000,5.960000] ordering_customer_name: [5.75.5.96] ordering_institution_bic: [CALAMANDRE] VALTER,TUMABARELLI MARTIN NAELI,INA ordering_institution_name: [BPPITRRDOXX,CRPIT2P] source: [DEFA_BITC03D960A03F4162BADEA09625D0,SEPA_31BB3A34622346FB3EBC33CF63C4B1]
ref	MUN_NOTPROVIDED_103_1408_UJT_NULL	0	10	11	> bic_1: [BBAESMM,CRAIT2R,DEUTDEF,PASCITMM,RVUALTZV] sender_bic: [BBAESMM,CRAIT2R,DEUTDEF,PASCITMM,RVUALTZV] session_sequence: [36e7aa2-16a8-4582-...
ref	MUN_NOTPROVIDED_103_1508_UJT_NULL	0	16	17	> bic_1: [BBRUBEC,ICRAITRR,SUNITMM,RVUALTZV] sender_bic: [BBRUBEC,ICRAITRR,SUNITMM,RVUALTZV] session_sequence: [00de965d-9dda-4412-a269-42b97a834b8,0cf...
ref	MUN_NOTPROVIDED_103_0812_UJT_NULL	0	578	579	> bic_1: [BACIT22,BAPIT22,BITATRR,BPMOT22,BPPITRR,CIPBITMM,CACIDEDDOXX,CRPIT2P,CRAIT2P,GEBAEBB,ICRAITRR,PASCITMM,RVUALTZV,SELBIT22,UBSWCHZ,UNCL...
ref	MUN_NOT_PROVIDED_103_0812_UJT_NULL	0	170	171	> bic_1: [BPPITRR,CIPBITMM,SUNITMM] sender_bic: [BPPITRR,CIPBITMM,SUNITMM] session_sequence: [02b8085-7527-4a85-a3da-a375e117c6b,034ed58-6f76-4395-a0e9-...
ref	FRK_NOTPROVIDED_103_2012_L_DE_NULL	0	1	2	> bic_1: [AARBDESXXXX,FMBKDEMMXXX] sender_bic: [AARBDESXXXX,FMBKDEMMXXX] session_sequence: [bc08b81-b3be-48df-a28c-079034ee198,9202467f-ce7a-457e-ba7...

- **Duplicados exactos:** filas idénticas.
- **Duplicados con variación:** el identificador coincide, pero alguna(s) columna(s) difiere(n). Es habitual cuando algún campo usado como identificador viene **NULL**.

La columna **variaciones** muestra el detalle de la diferencia por columna, p. ej.:
bic_1: [valor_1, valor_2] (y así sucesivamente para el resto de campos afectados).

		^A _C exact_duplicates	^A _C dupes_w_variatio...	^A _C occurrences	^A _C variations
1	PROVIDED_103_08:12_I_IT_NULL	0	170	171	> bic_1: [BPPIITRR,CIPBITMM,ISUN
2	PROVIDED_103_08:12_I_IT_NULL	0	578	579	> bic_1: [BACRIT22,BAPPIT22,BITA
3	_NULL_06:14_I_IT_SCL	0	1	2	v bic_1: [BPPIITRR,CRPPIT2P] sen [BPPIITRRXXXX,CRPPIT2P] trans ordering_institution_bic: [CALAM [SEPA_881C03D960A83F41B28A

3) Definición de la clave compuesta (ejecuciones mostradas)

Se utilizó la siguiente **clave compuesta**:

geo, uetr, type, time, in_out, segregation, sender_reference

Ejemplo de identificador (formado con la clave anterior):

BR_0007589b-71e8-4d5c-9f69-134e9037631a_103_09:14_0_3565729148

4) Opciones y comportamiento

- **checkDuplicates:** si está **desactivado**, la evaluación de duplicados se omite y results_duplicates queda **vacía incluso si existen duplicados**.
- **includeEqualsInDiff:** si está **activado**, results_differences incluirá también los **EXACT_MATCH** y los **MATCH** de los **PARTIAL MATCH**. Si está **desactivado**, esa parte no se escribe y la tabla puede quedar vacía (en igualdad total).

5) Configuración actual en Variable de Airflow (PRO)

- **Tablas a comparar:** Se definen refTable (tabla de referencia) y newTable (tabla nueva).
- **Rangos de ventanas independientes por tabla,** refWindowDays y newWindowDays. ejemplos:

- "refWindowDays": "0..0", Date actual
- "newWindowDays": "-1..0", Desde d actual hasta d-1
- "newWindowDays": "-1..+1", Desde d actual hasta d-1 y d+1
- "newWindowDays": "0..+2", Desde d actual hasta d+2

Todas las posibilidades asc y desc para comparar ventanas independientes por tabla.

- **Clave compuesta (compositeKeyCols):** Define los campos que forman el identificador único.
- **Columnas a ignorar (ignoreCols):** Campos que no se tienen en cuenta en la comparación.
- **Opciones funcionales:**
 - checkDuplicates: Indica si se debe evaluar la existencia de duplicados (establecido como true en la configuración actual).
 - includeEqualsInDiff: Indica si se deben incluir los EXACT_MATCH (coincidencias perfectas) en la tabla de diferencias.

```
AMLTTOOLS_tables_comparator_swift {  
  "refTable": "scib_cm_cmplnc_trans_messages_swift_s3.swift_transactions",  
  "newTable": "scib_bu_cmplnc_trans_messages_swift_s3.swift_transactions_new",  
  
  "initiativeName": "swift",  
  "tablePrefix": "scib_bu_cmplnc_trans_messages_swift_s3.results_",  
  "outputBucket": "s3a://scib-pro-bu-aml/internal-tools",  
  "partitionSpec": "geo=*/data_date_part=YYYY-MM-dd/",  
  
  "refWindowDays": "0..0",  
  "newWindowDays": "-1..0",  
  "compositeKeyCols": ["geo", "uetr"],  
  "ignoreCols": ["original_message", "data_timestamp_part"],  
  "checkDuplicates": true,  
  "includeEqualsInDiff": true  
}
```

Observación: En el ejemplo de ejecuciones se amplió la clave efectiva (*geo*, *uetr*, *type*, *time*, *in_out*, *segregation*, *sender_reference*). Ajusta `compositeKeyCols` cuando necesites esa granularidad.

6) Configuración utilizada en Postman (ejemplo)

POST: https://hybridbalancer.scib.gs.corp/v2/execution/aml_internal_tools

```
{
  "preferences": ["azure_databricks_v2", "cloudera"],
  "size": "S",
  "parameters": [
    "refTable=scib_cm_cmplnc_trans_messages_swift_s3.swift_transactions",
    "newTable=scib_cm_cmplnc_trans_messages_swift_s3.swift_transactions",
    "initiativeName=swift",
    "tablePrefix=scib_bu_cmplnc_trans_messages_swift_s3.results_",
    "outputBucket=s3a://scib-pro-bu-aml/internal-tools",
    "executionDate=2025-09-30",
    "partitionSpec=data_date_part=YYYY-MM-dd/",
    "refWindowDays=0..0",
    "newWindowDays=-1..0",
    "compositeKeyCols=geo,uetr,type,time,in_out,segregation,sender_reference",
    "ignoreCols=original_message,data_timestamp_part",
    "checkDuplicates=true",
    "includeEqualsInDiff=true"
  ]
}
```

7) Lectura rápida / Checklist

1. Mira **results_summary** para el estado general: Unique IDs, Total rows, Total diff, Global Quality, y usa **Samples por métrica (hasta 6 IDs)** para pinchar ejemplos al vuelo en **results_differences** o **results_duplicates**.
2. Abre **results_differences** para el **detalle de diferencias** (y exactos si los guardas).

3. Revisa **results_duplicates** para ver **dónde** y **cómo** se duplican los identificadores (both / duplica REF / duplica NEW, exactos o con variación).
4. Recuerda que **duplicados afectan a Global Quality** y que **las opciones de configuración** pueden hacer que algunas tablas aparezcan vacías intencionalmente.

8) Glosario rápido

- **REF / NEW:** tabla de referencia / tabla nueva a comparar.
- **EXACT_MATCH:** coincidencia 1:1 perfecta (todas las columnas).
- **PARTIAL_MATCH:** coincidencia 1:1 con diferencias en alguna columna.
- **GAP:** elementos presentes en una tabla y ausentes en la otra (se listan ejemplos).
- **Duplicados exactos:** filas 100% idénticas.
- **Duplicados con variación:** mismo identificador con diferencias en otras columnas.
- **Global Quality:** métrica global que no llega a 100% si hay duplicados.

9) Job Control-M

B	C	D	E	F
ODATE	FOLDER_NAME	JOB_NAME	FORCE_IGNOREING_CALENDAR	FECHA_EJECUCION
20250919	JACP-P-SWIFT-800036032	P-SWIFT_INTERNAL_COMPARISON-D	YES	ASAP
			✓	
			✓	
			✓	
			✓	
			✓	

Fin