

MOSTRAR 2 CASOS DE EJECUCION EN PRO COMPARANDO LA TABLA REF CONSIGO MISMA.

CASO 1: IGUALDAD PERFECTA: 2025-09-21

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	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_008
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%	-

Tabla Diferencias (bd.results_diferencias)

*Nota: Veríamos los **EXACT_MATCH** en caso de estar activa la opción de mostrarlos, en caso contrario esta tabla estaría vacía.

		column	value_ref	value_n...	results
1	i626492-470b-46c8-afa7-552194f657d6_103_09:29_I_BR_MBB17D386535C0A	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
2	7a2331-2564-4261-bceb-f89b5d157be2_103_10:49_I_BR_MBB4F727950CE...	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
3	8fc9b6-c2cc-41bc-92e2-35d776c3106f_103_15:04_I_BR_MBB81ECE45A2181	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
4	9989e4-1a76-4b04-90c6-79582a38016a_103_08:46_I_ES_1560696010018553	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
5	590410-919b-4933-8b78-96a88af7fbbd_103_16:06_O_ES_2025092200079985	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
6	K_ae7149ef-7fce-4859-ae91-95353962aed5_103_18:45_O_MX_49986960100...	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
7	db065c-263a-493b-8d80-bbdb0dd16fee_103_05:08_I_PL_2509210856947686	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
8	0fd9e0-a14f-4922-ae72-fce50be83ae1_103_23:09_I_PL_2509210868376233	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
9	0f11eb-58ba-4b56-a6f8-56880137d106_103_18:48_I_ES_2354696010056626	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
10	0d9313-f95a-488c-a50e-ff6c96c9644b_103_16:03_I_ES_0950696010011159	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
11	ifdfbf-ba00-4a22-b2e9-5ffb3199fa5e_103_12:11_I_ES_1736696010045921	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
12	d6d32c-cb56-4eec-927c-3a4eff766241_103_03:49_O_ES_MBB2B1A27CB9CC9	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
13	1f8749c-4c5e-4144-bdb0-5c0e6c45c12a_103_12:18_O_MX_D0452650026901	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
14	9ff9147-98a6-4169-af1b-eafb35b6d782_103_10:46_I_MX_OPAY25092100027	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH
15	c774a2e-1bab-4460-bea2-c41c01f4af0f_103_22:55_I_MX_OPAY25092000141	ALL_COLUMNS	ALL_MATCH	ALL_MATCH	EXACT_MATCH

Tabla Duplicados (bd.results_duplicates)

Esta tabla estaría vacía.

*Nota: también tenemos una opción para desactivar la evaluación de los duplicados de tal manera que aunque existiesen los ignoramos y esta tabla se queda vacía.

CASO 2: MAS TIPICO Y AL PARECER CORRECTO FUCNCIONALMENTE: 2025-09-24

Tabla Ejecutiva (bd.results_summary)

- **Duplicados exactos:** Filas idénticas
- **Duplicados con variación:** El identificador que hayamos decidido funcionalmente hace match como valor único, pero algunas del resto de sus columnas adquieren valores diferentes. En este caso suele pasar cuando alguno de los campos que se usa como identificador viene a NULL.
- **Variaciones:** En esta columna podemos ver el detalle de la duplicidad.
*En el siguiente ejemplo vemos que **exact_duplicates** está 0, no hay duplicados idénticos, pero hay 2 ocurrencias por identificados una de ellas será la que debería ser la buena y la que está en **dupes_w_variation** tiene un 1 nos indica que hay un duplicado con variación.*
 En la columna variaciones podemos ver el detalle del duplicado, para este ejemplo miramos el primero que se duplica columna: **bic_1**: [valor 1, valor 2] y así con el resto podremos ver directamente que es lo que se duplica y como.

		exact_duplicates	dupes_w_variatio...	occurrences	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	✓ bic_1: [BPPIITRR,CRPPIT2P] sen [BPPIITRRXXX,CRPPIT2P] trans ordering_institution_bic: [CALAN [SEPA_881C03D960A83F41B28A

Otras aclaraciones:

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%

Bloque 1: es información directa y general, no escribimos detalle solo vemos los count.

- **Unique IDS:** Total de ocurrencias únicas según el id que hayamos decidido.
- **Total rows :** Cuenta el total de filas que llegan (cuenta duplicados)
- **Total diff:** siempre es una resta de filas de (NEW – REF).
- **Global Quality:** Tiene en cuenta en su fórmula para dar el 100% que no haya duplicados debido a que es una variable de su fórmula.

Bloque 2: El detalle de las métricas está en la tabla: **bd_results_differences**

- **EXACT_MATCH (1:1):** Se ha conseguido el objetivo de que todas las columnas sean idénticas.
- **PARTIAL_MACTH (1:1)** Hace Match el identificador que hayamos elegido, pero hay diferencias entre la tabla de referencia y nueva en algunas de sus columnas. Veremos 2 categorías en la tabla de diferencias los **MATCH** y **NO_MATCH**, en caso de estar habilitada la opción de guardar también los MATCH.
- **GAP:** Lo que tiene la una que no tiene la otra. Tendremos siempre 6 ejemplos de identificadores de muestra.

Bloque 3: El detalle de estas métricas se escribe en la tabla: **bd_results_duplicates**

- **DUPS:** Tenemos 3 tipos
 1. **Duplicados en ambas tablas,** REF y NEW duplican el mismo identificador. En caso de que ambos dupliquen se resta de la cuenta de duplicados independientes de cada origen.
 2. **Duplica REF**
 3. **Duplica NEW**

La clave que se ha utilizado para estas ejecuciones:

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

Ejemplo de cómo queda un id en tabla diferencias y duplicados con la definición de clave anterior:

BR_0007589b-71e8-4d5c-9f69-134e9037631a_103_09:14_O_3565729148

Configuración actual en variable de Airflow PRO:

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

Configuración postman:

```
{
  "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=2025-09-25",
    "partitionSpec=data_date_part=YYYY-MM-dd/",
    "compositeKeyCols=geo,uetr,type,time,in_out,segregation,sender_reference",
    "ignoreCols=original_message,data_timestamp_part",
    "checkDuplicates=true",
    "includeEqualsInDiff=true"
  ]
}
```