

# **Issue with Distributed joins**

## **Problem statement:**

We are trying to join three tables for two cases.

- case1: Joining columns are non-primary keys.
- case2: Joining columns are primary keys.
- For case2 distributed joins are not giving right results.

Table1: person

NAME	AGE	CITY_ID
Hope	27	1.2
Madhavi	23	3.2
Shravya	25	1.1
Deeps	25	1.2
Harika	26	2.4
Kiran	26	1.1
Srinivas	24	3.2

tableName: person\_P  
cache:Partitioned  
primaryKey: name

Table2: medical\_info

NAME	AGE	BLOOD_GROUP
Diggi	27	B+
Kiran	26	O+
Harika	26	AB+
Madhavi	23	A+

tableName: medical\_info\_P  
cache:Partitioned  
primaryKey: id

Table3: blood\_group\_info\_P

BLOOD_GROUP	UNIVERSAL_DONOR
A+	A+AB+
AB+	AB+
B+	B+AB+
O+	O+A+B+AB+

tableName: blood\_group\_info\_P  
cache:Partitioned  
primaryKey: blood\_group

Table4: blood\_group\_info\_PJ

BLOOD_GROUP	UNIVERSAL_DONOR
A+	A+AB+
AB+	AB+
B+	B+AB+
O+	O+A+B+AB+

tableName: blood\_group\_info\_PJ  
cache:Partitioned  
primaryKey: id

## CASE 1: SQL on table1, table2, table4:

Allow non-collocated joins ✓

```
SELECT person.id, person.name, medical_info.blood_group ,blood_group_info_P.universal_donor
FROM person
LEFT JOIN medical_info ON medical_info.name = person.name
LEFT JOIN blood_group_info_PJ ON blood_group_info.blood_group = medical_info_P.blood_group
```

The screenshot shows a SQL query execution interface. The query is as follows:

```
1 select person PJ.id, person PJ.name, medical_info pj.blood_group, blood_group_info_PJ.universal_donor from person_PJ
2 left join medical_info_PJ on person PJ.name = medical_info PJ.name
3 left join blood_group_info_PJ on medical_info PJ.blood_group = blood_group_info_PJ.blood_group
4
```

Execution controls include buttons for "Execute", "Execute on selected node", and "Explain". The interface also shows "Rows per page: 100" and "Max pages: Unlimited".

On the right, there is a "Caches" section with a search bar and a list of caches. The "Allow non-collocated joins" checkbox is checked.

The results table is displayed below the query, showing 7 rows. The columns are ID, NAME, BLOOD\_GROUP, and UNIVERSAL\_DONOR.

ID	NAME	BLOOD_GROUP	UNIVERSAL_DONOR
1001	Shravya		
1007			
1006	Deeps		
1003	Harika	AB+	
1002	Kiran	O+	O+A+B+AB+
1005	Madhavi		
1004	Srinivas		

At the bottom left, there is a "Show query" link.

Please find the sql result details in next slide:

Summary :

## CASE-1 Results: Correct and as expected

```
SELECT
```

```
  __Z0.ID AS __C0_0,
```

```
  __Z0.NAME AS __C0_1,
```

```
  __Z1.BLOOD_GROUP AS __C0_2,
```

```
  __Z2.UNIVERSAL_DONOR AS __C0_3
```

```
FROM PUBLIC.PERSON__Z0
```

```
  /* PUBLIC.PERSON_NAME_ASC_IDX_proxy */
```

```
LEFT OUTER JOIN PUBLIC.MEDICAL_INFO __Z1
```

```
  /* batched:broadcast PUBLIC.MEDICAL_INFO_NAME_ASC_IDX: NAME = __Z0.NAME */
```

```
  ON __Z0.NAME = __Z1.NAME
```

```
LEFT OUTER JOIN PUBLIC.BLOOD_GROUP_INFO_PJ __Z2
```

```
  /* batched:broadcast PUBLIC.BLOOD_GROUP_INFO_PJ_BLOOD_GROUP_ASC_IDX: BLOOD_GROUP =  
  __Z1.BLOOD_GROUP */
```

```
  ON __Z1.BLOOD_GROUP = __Z2.BLOOD_GROUP
```

## CASE 2: SQL on table1, table2, table3:

Allow non-collocated joins ✓

```
SELECT person.id, person.name, medical_info.blood_group ,blood_group_info_P.universal_donor
FROM person
LEFT JOIN medical_info ON medical_info.name = person.name
LEFT JOIN blood_group_info_P ON blood_group_info.blood_group = medical_info_P.blood_group
```

The screenshot shows a SQL query execution interface. The query is as follows:

```
1 SELECT person.P.id, person.P.name, medical_info.P.blood_group ,blood_group_info_P.universal_donor FROM person_P
2 left join medical_info_P on medical_info_P.name = person_P.name
3 left join blood_group_info_P on blood_group_info_P.blood_group = medical_info_P.blood_group
4
```

Execution controls include buttons for "Execute", "Execute on selected node", and "Explain". The "Rows per page" is set to 100, and "Max pages" is Unlimited. The "Allow non-collocated joins" checkbox is checked.

The results table shows the following data:

ID	NAME	BLOOD_GROUP	UNIVERSAL_DONOR
2004	Srinivas		
2006	Deeps		
2001	Shravya		
2007	Hope		

At the bottom left, there is a link to "Show query". At the bottom right, there is an "Export" button.

Please find the sql result for case2 details in next slide:

Summary :

## CASE-2 Results: In-correct ❌

SELECT

\_\_Z0.ID AS \_\_C0\_0,

\_\_Z0.NAME AS \_\_C0\_1,

\_\_Z1.BLOOD\_GROUP AS \_\_C0\_2,

\_\_Z2.UNIVERSAL\_DONOR AS \_\_C0\_3

FROM PUBLIC.PERSON \_\_Z0

/\* PUBLIC.PERSON\_ID\_ASC\_IDX\_proxy \*/

LEFT OUTER JOIN PUBLIC.MEDICAL\_INFO \_\_Z1

/\* batched:broadcast PUBLIC.MEDICAL\_INFO\_NAME\_ASC\_IDX: NAME = \_\_Z0.NAME \*/

ON \_\_Z0.NAME = \_\_Z1.NAME

LEFT OUTER JOIN PUBLIC.BLOOD\_GROUP\_INFO\_P \_\_Z2

**/\* batched:unicast PUBLIC.\_key\_PK\_proxy: BLOOD\_GROUP = \_\_Z1.BLOOD\_GROUP \*/**

ON \_\_Z1.BLOOD\_GROUP = \_\_Z2.BLOOD\_GROUP

## **Conclusion :**

Distributed joins in case2 is not giving right results.