

WWW.PICODATA.IO

SBROAD

Internal design

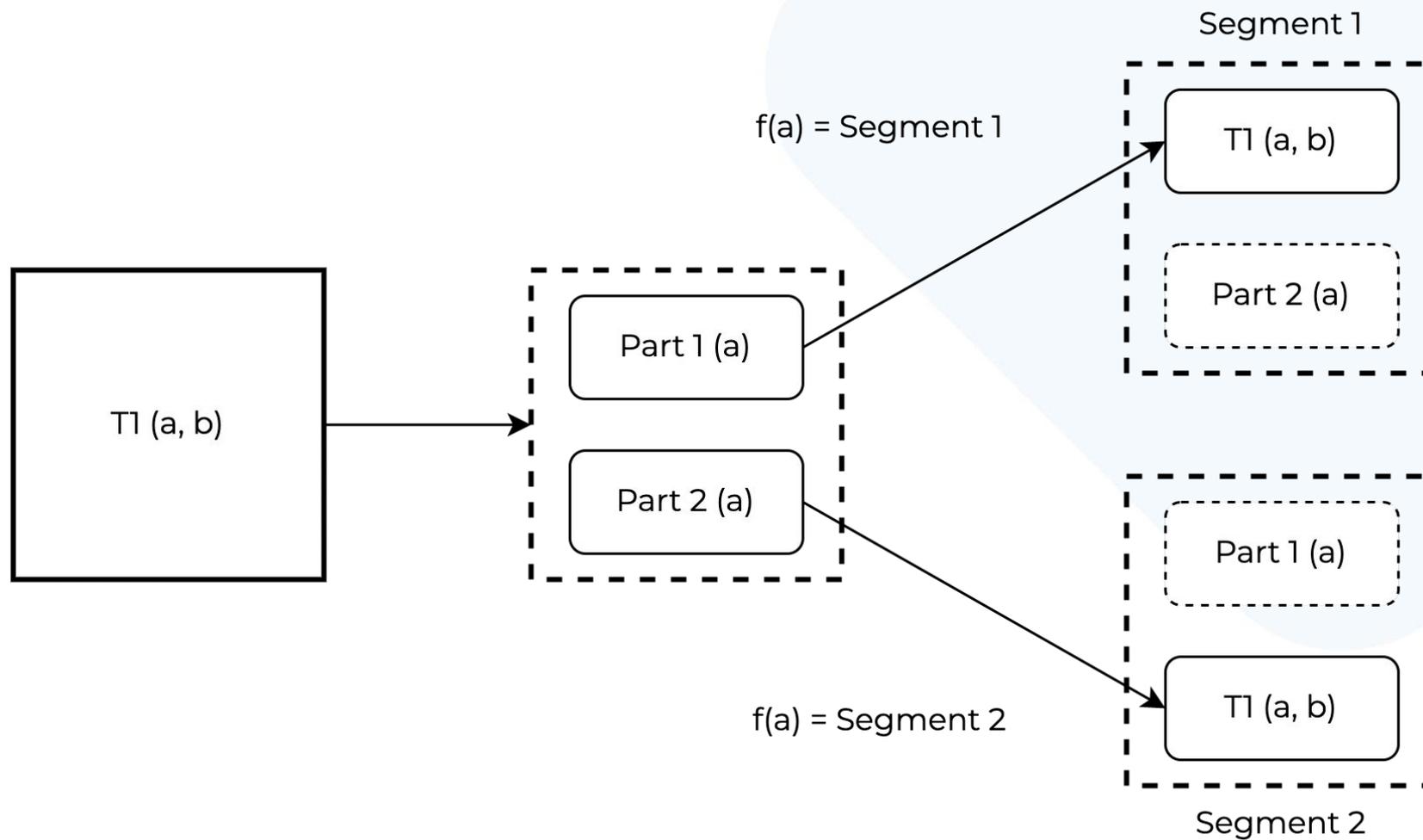


PICODATA

Distributed SQL

What is the problem?

Data segmentation

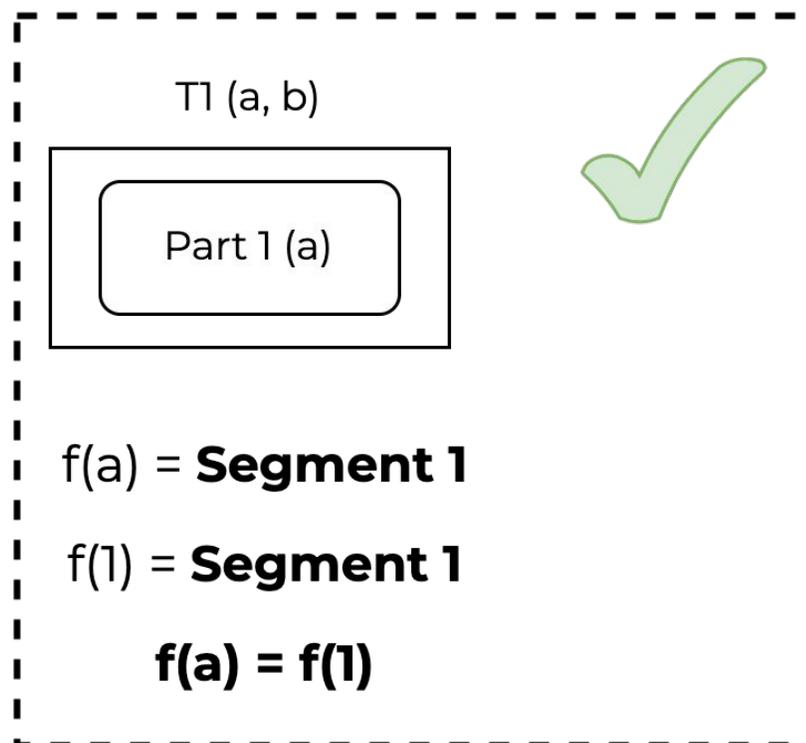


Distributed query

Where to execute?

SELECT * FROM t1 WHERE a = 1

Segment 1



Segment 2

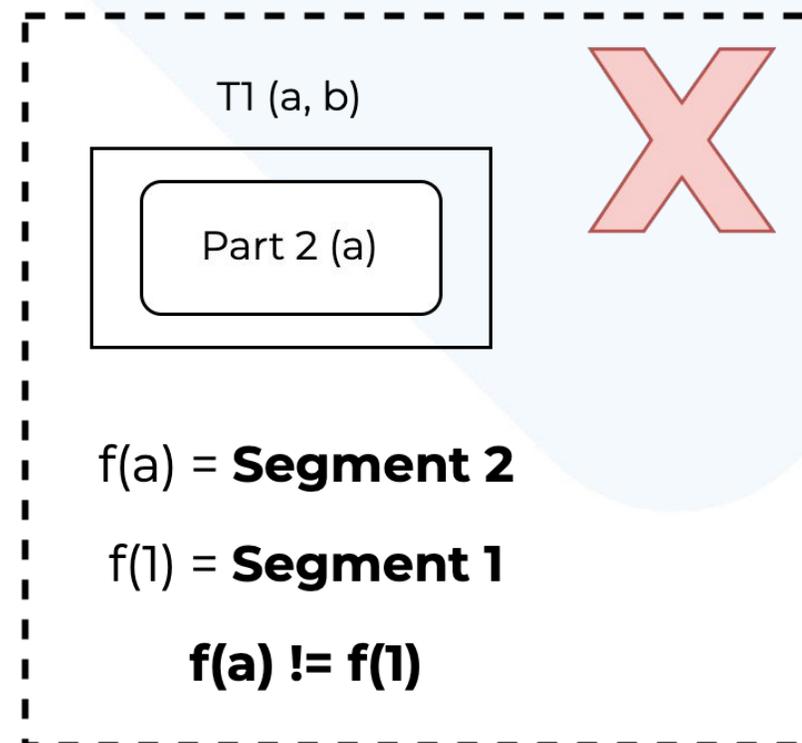
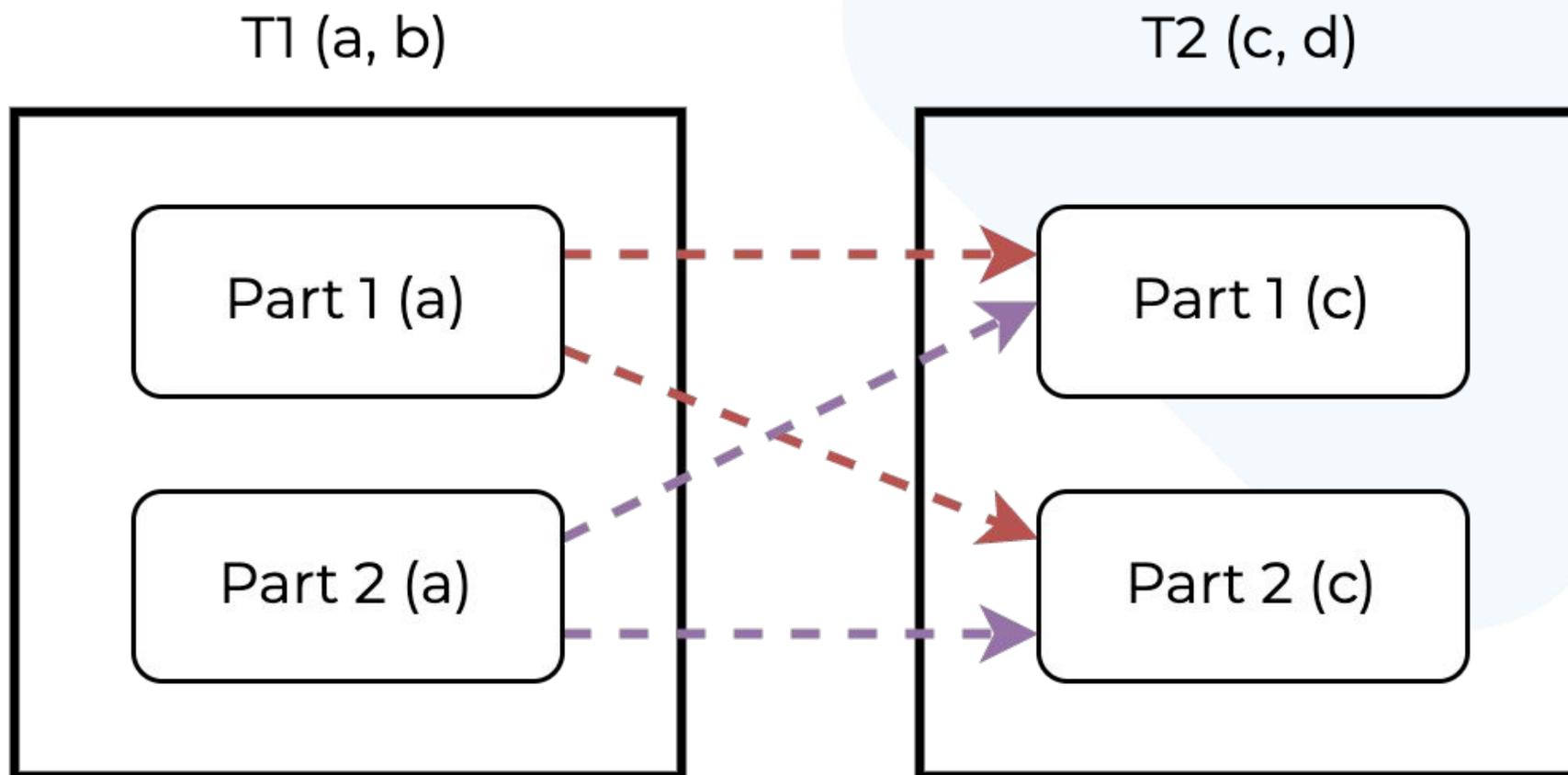
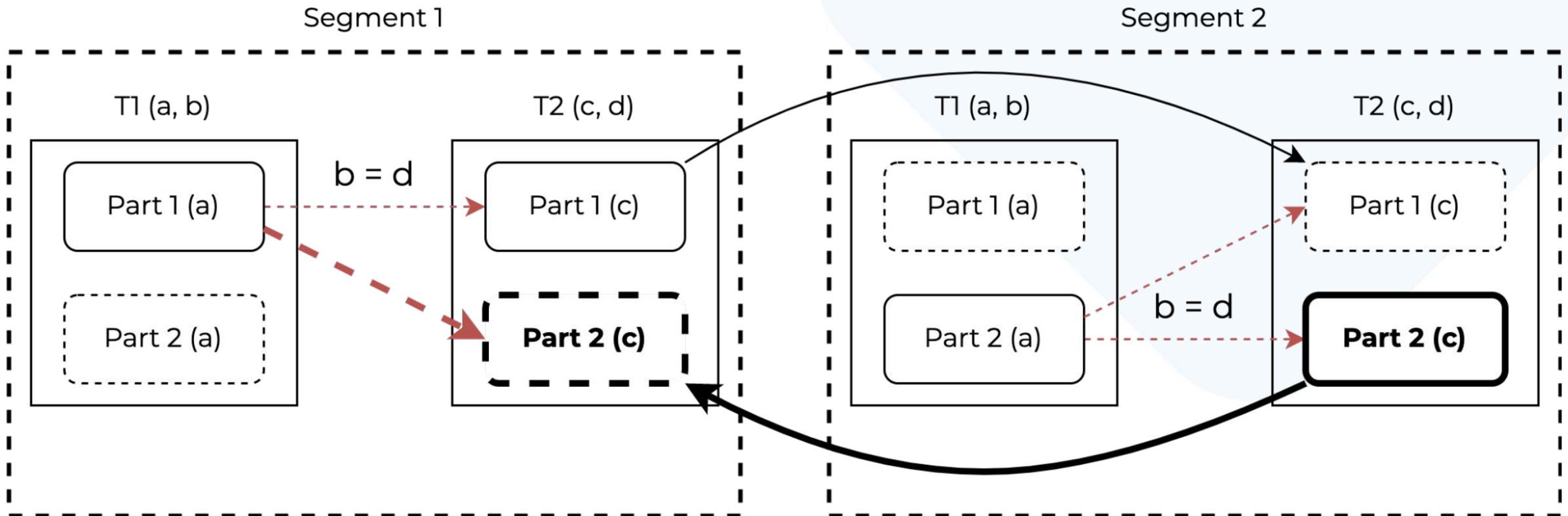


Table join



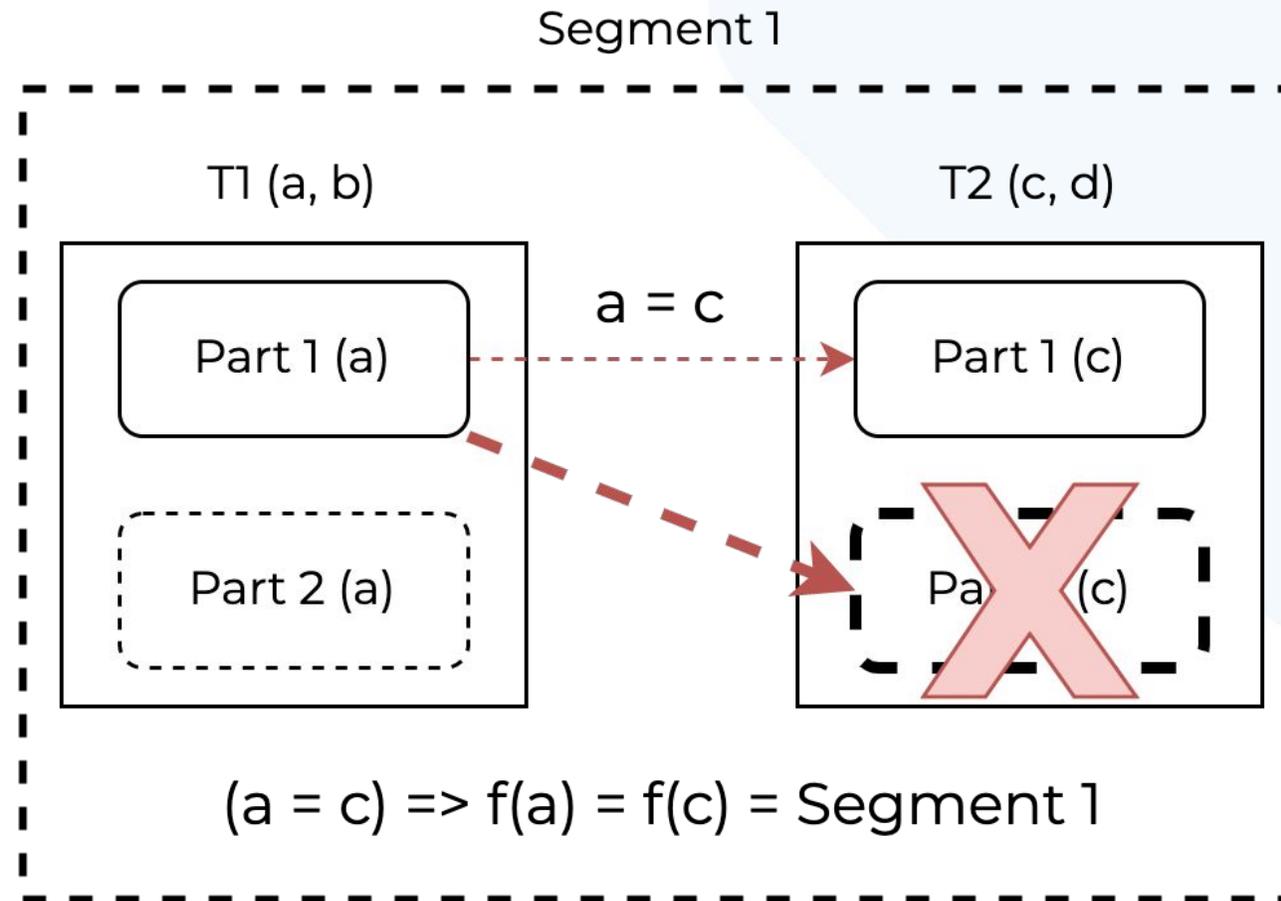
Distributed join

Full copy



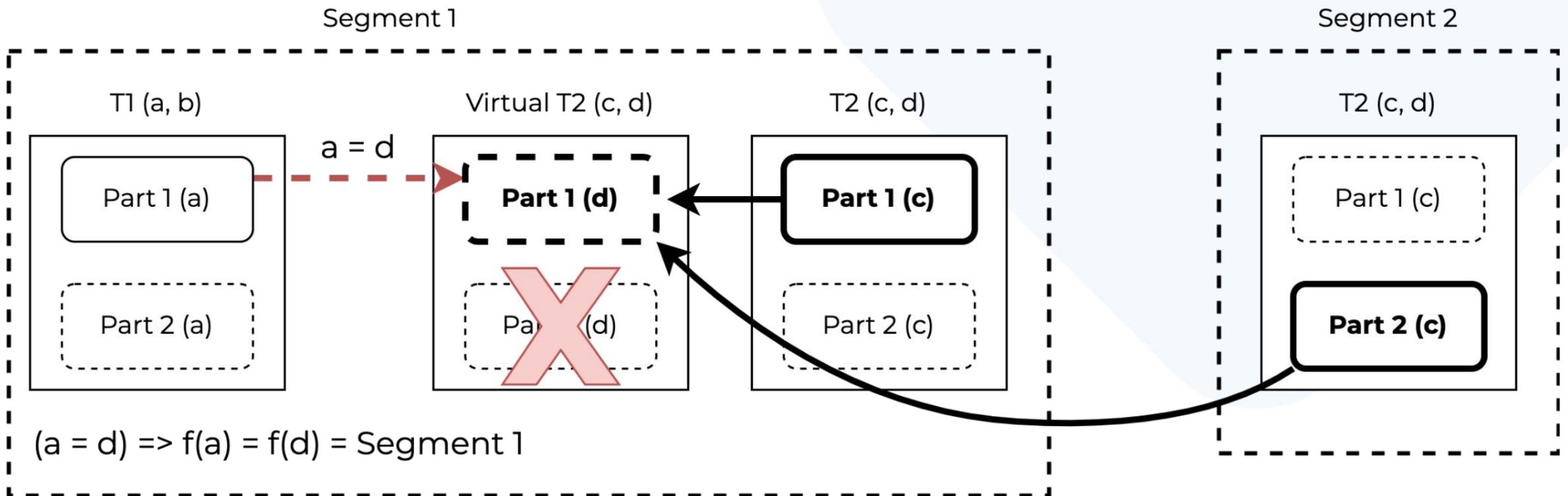
Distributed join

Local



Distributed join

New key



Distributed query

Data motion

Segment 1



Segment 2

Coordinator



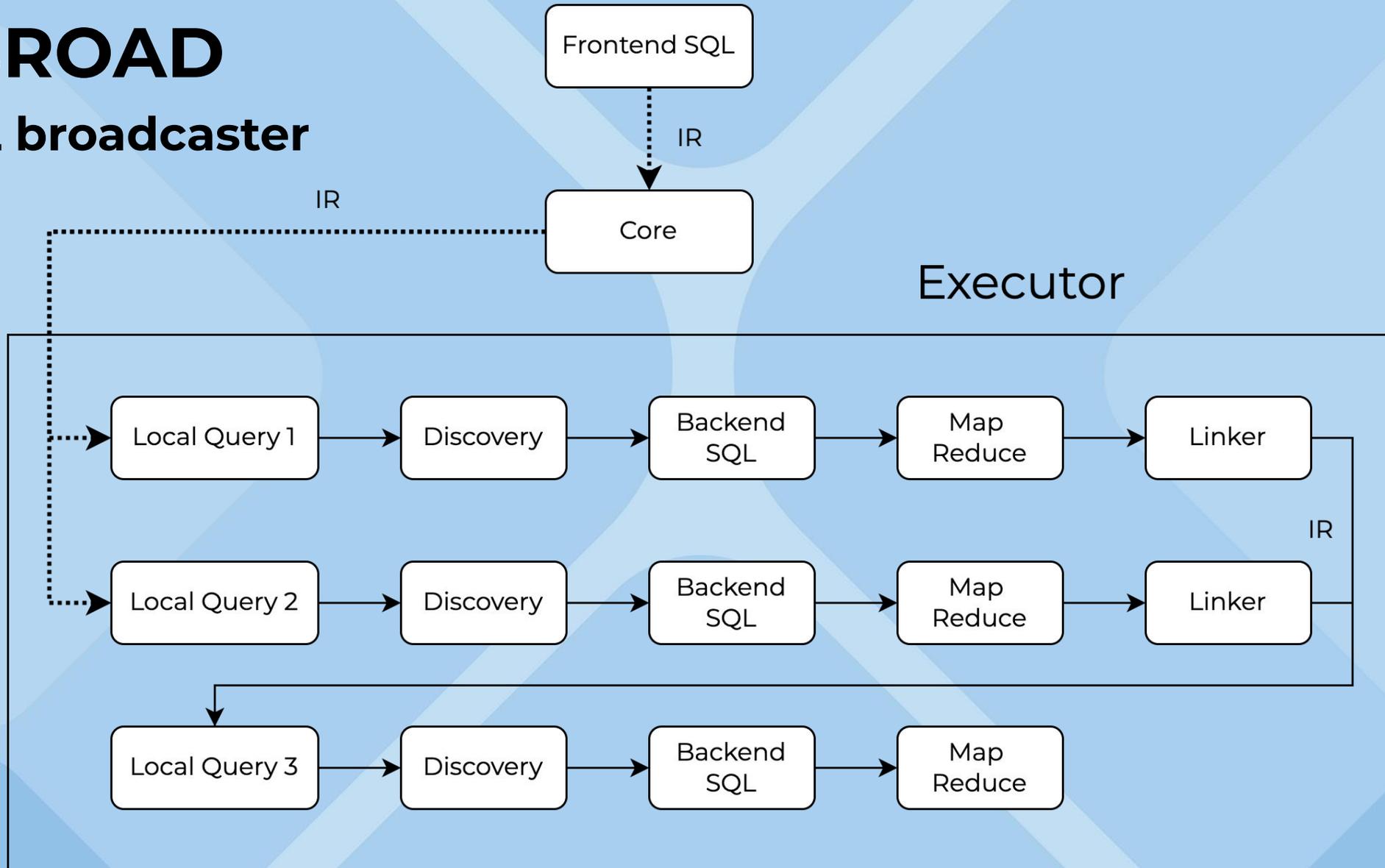
Segment 1



Segment 2

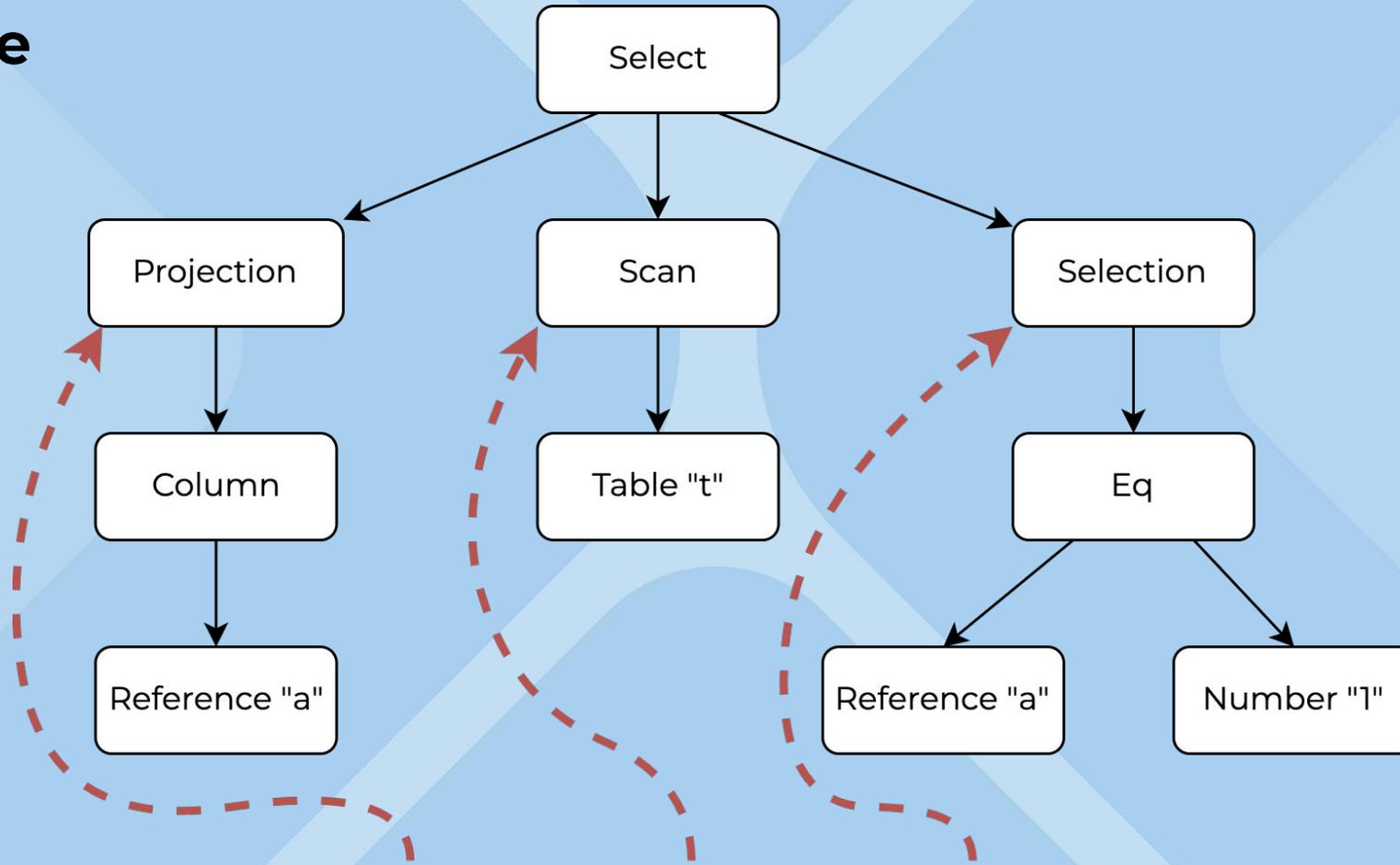
SBROAD

SQL broadcaster



Frontend SQL

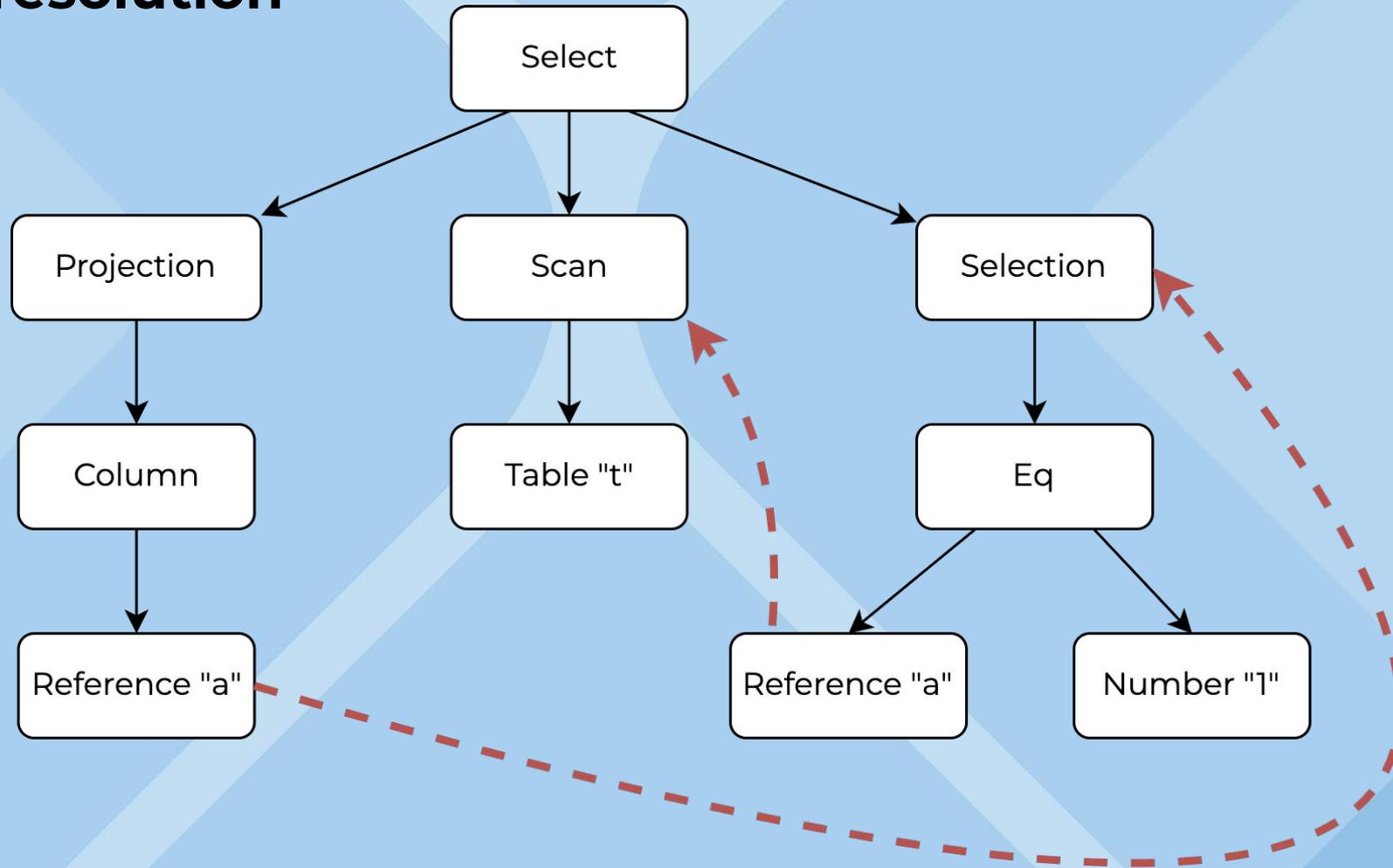
Parse tree



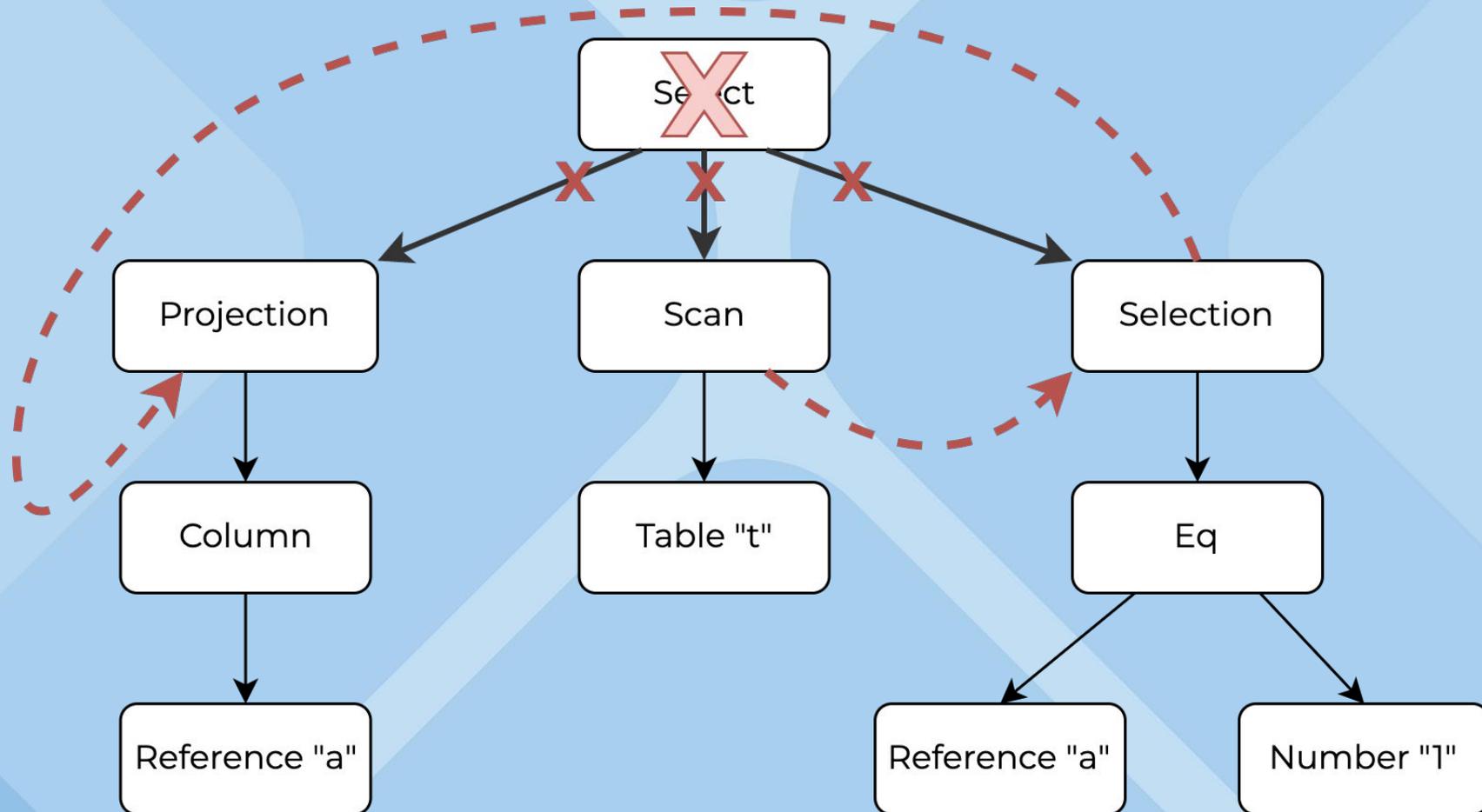
SELECT a FROM t WHERE a = 1

Frontend SQL

References resolution

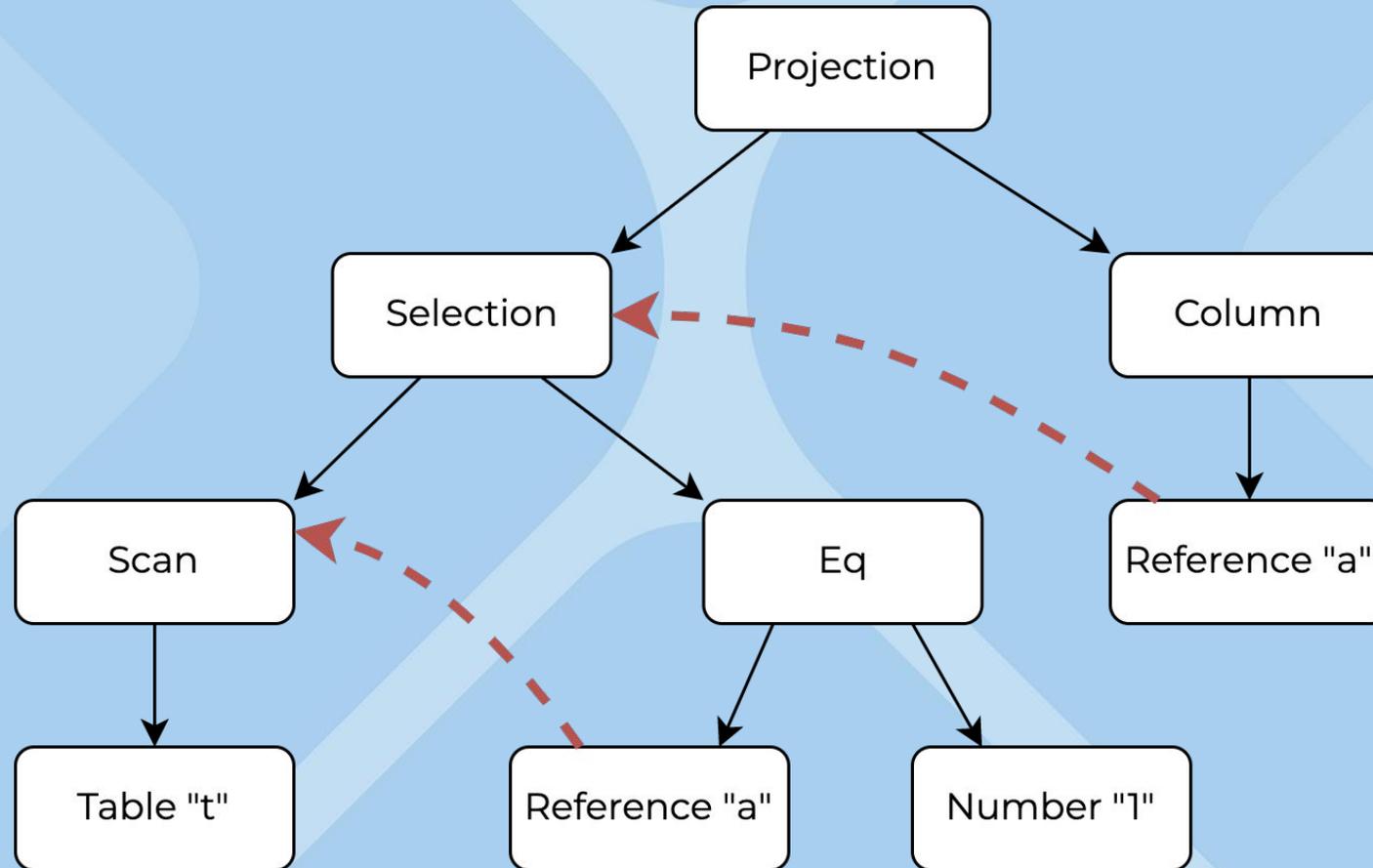


Frontend SQL Transformation



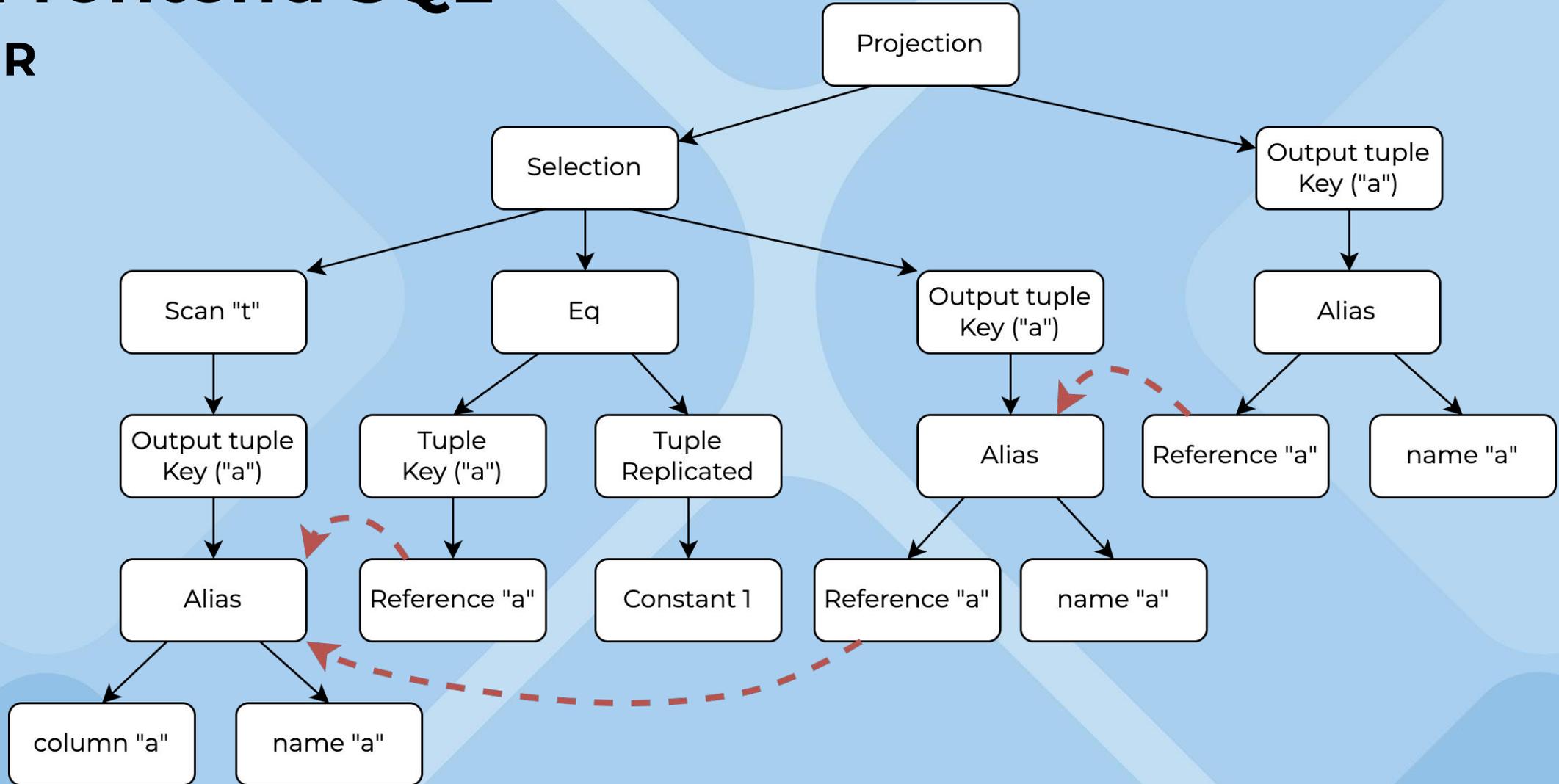
Frontend SQL

AST

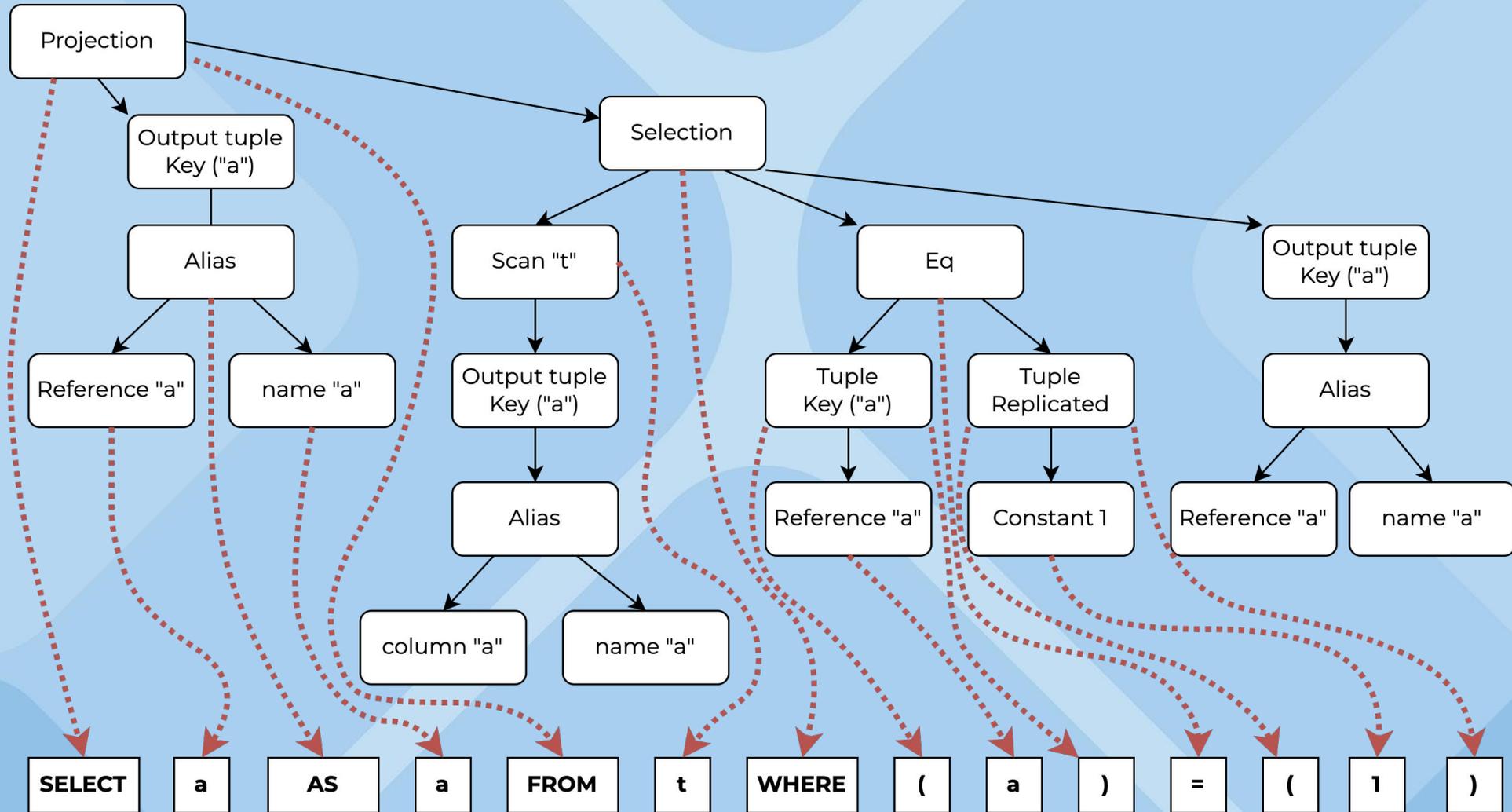


Frontend SQL

IR



Backend SQL





PICODATA

Core

Distribution conflict

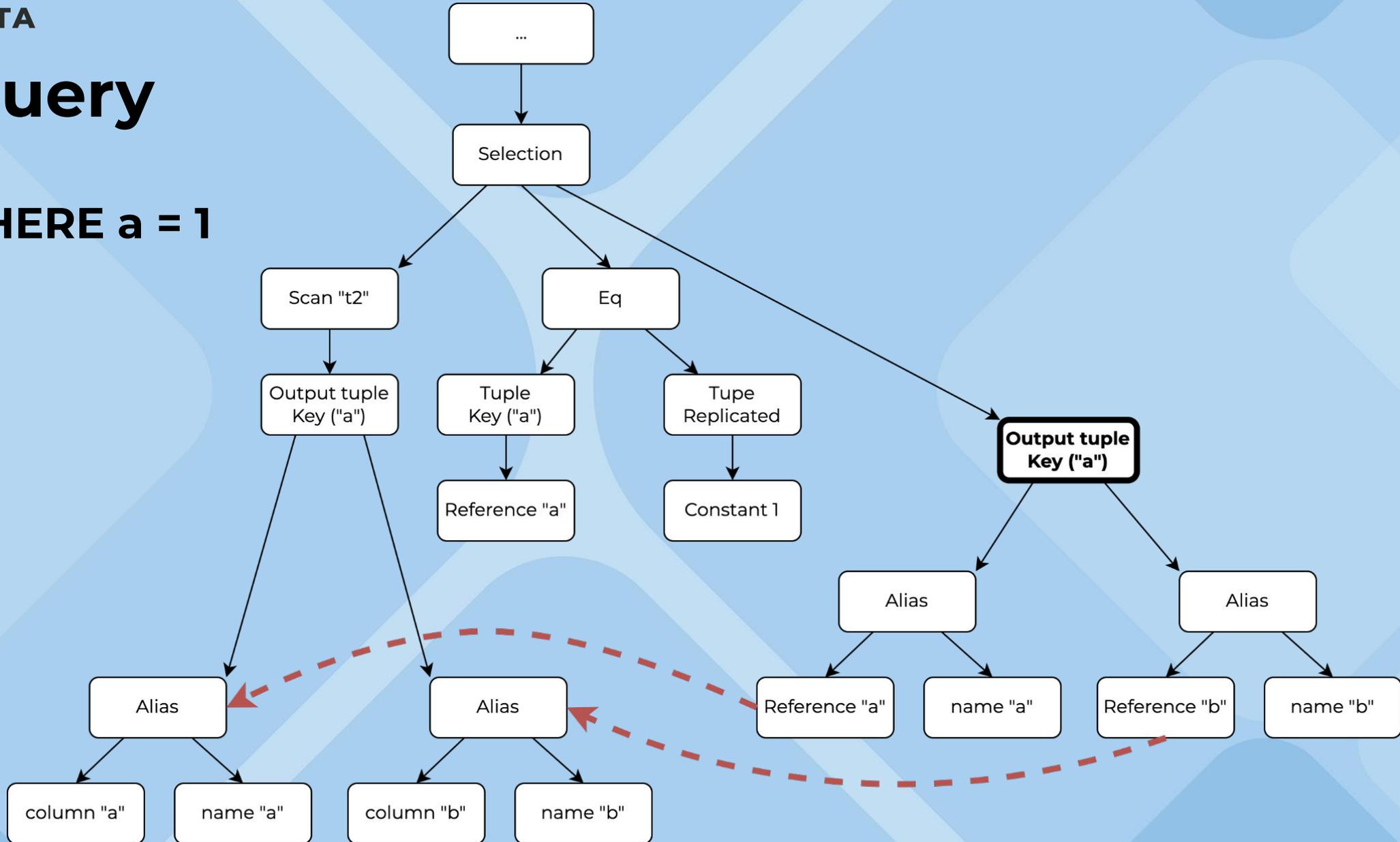
table t1 (a) distributed by (a)

table t2 (a, b) distributed by (a)

```
SELECT a FROM t1 WHERE a IN (SELECT b FROM t2 WHERE a = 1)
```

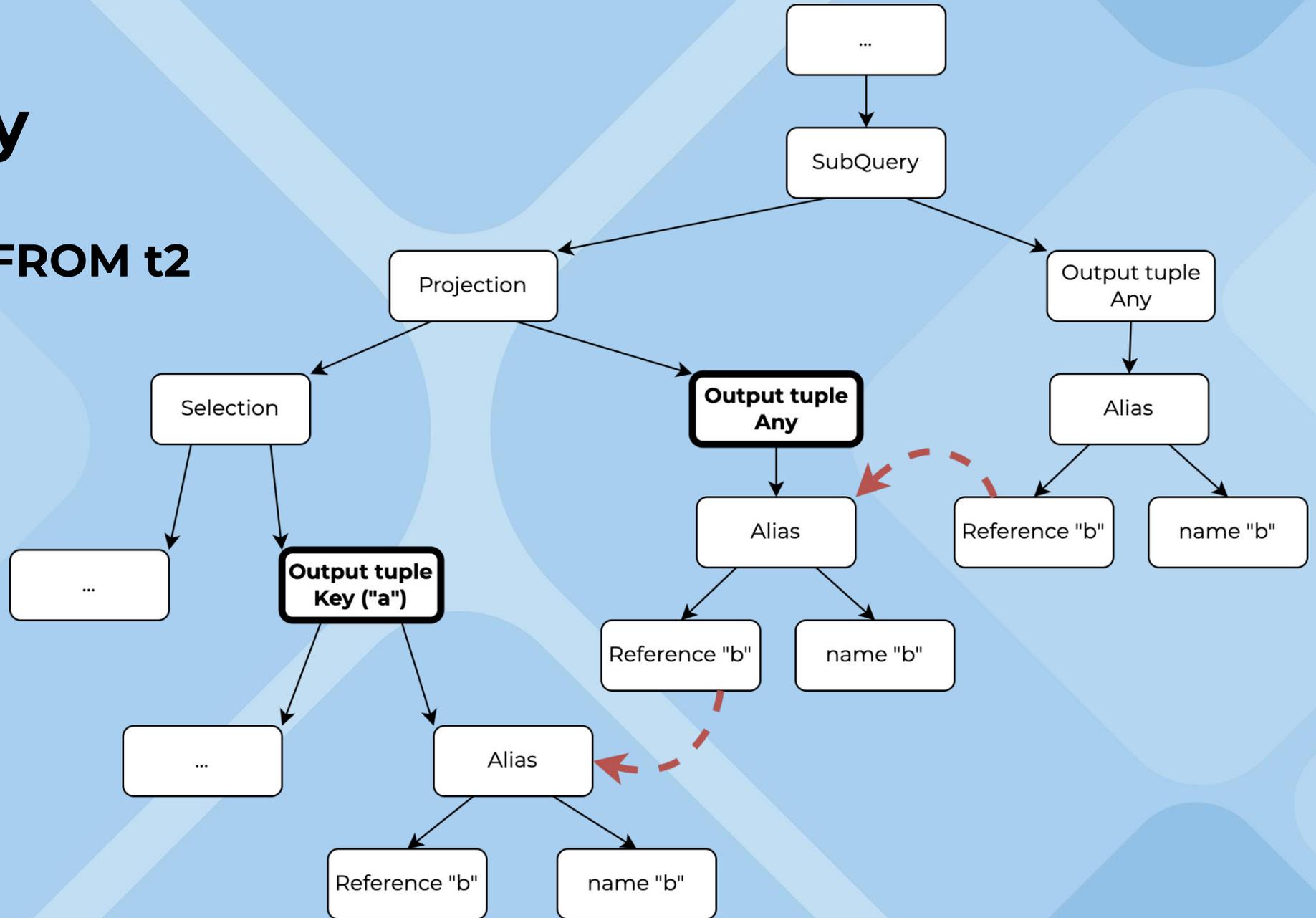
Subquery

... t2 WHERE a = 1



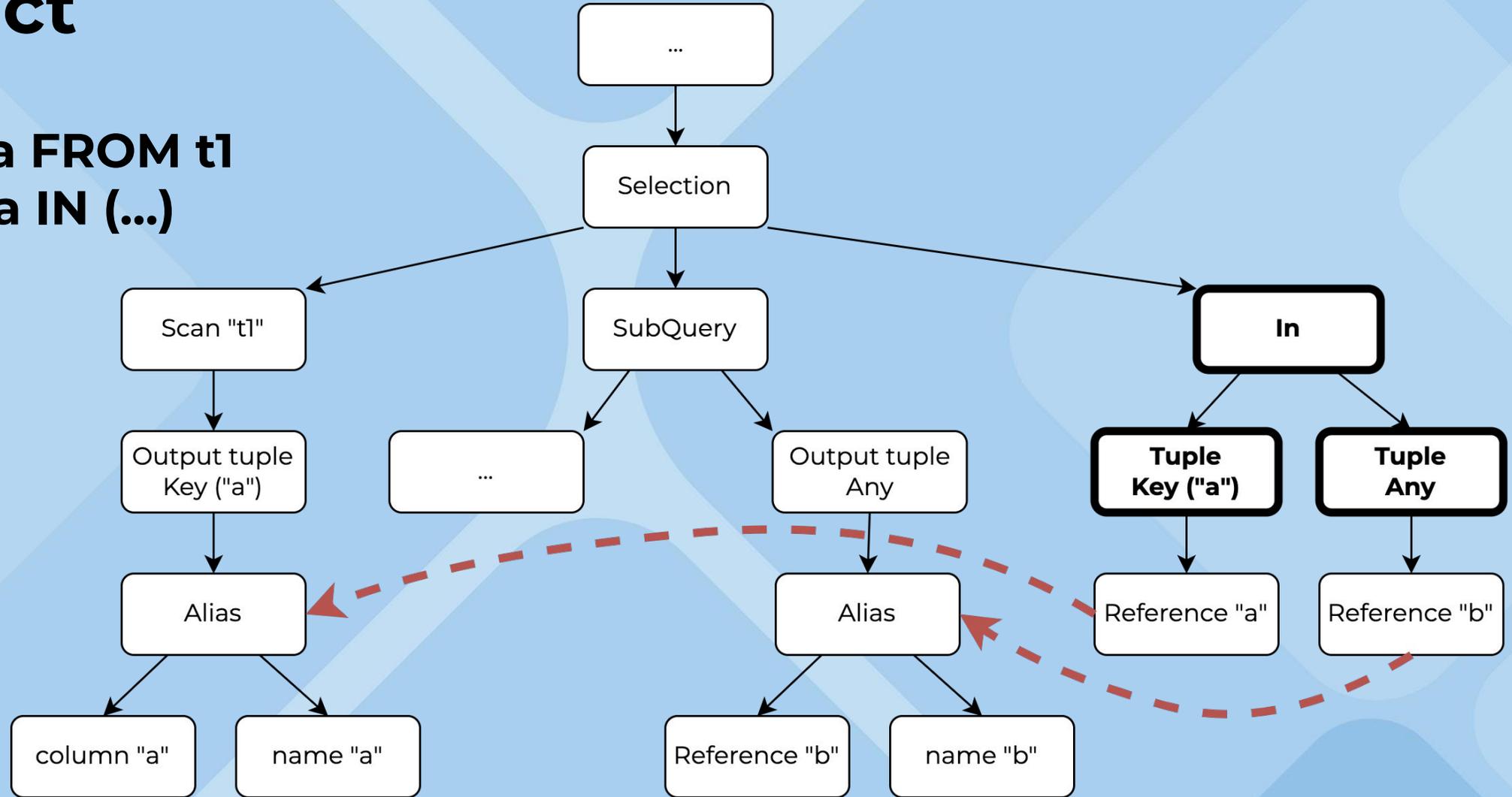
Subquery

... (SELECT b FROM t2
WHERE a = 1)



Conflict

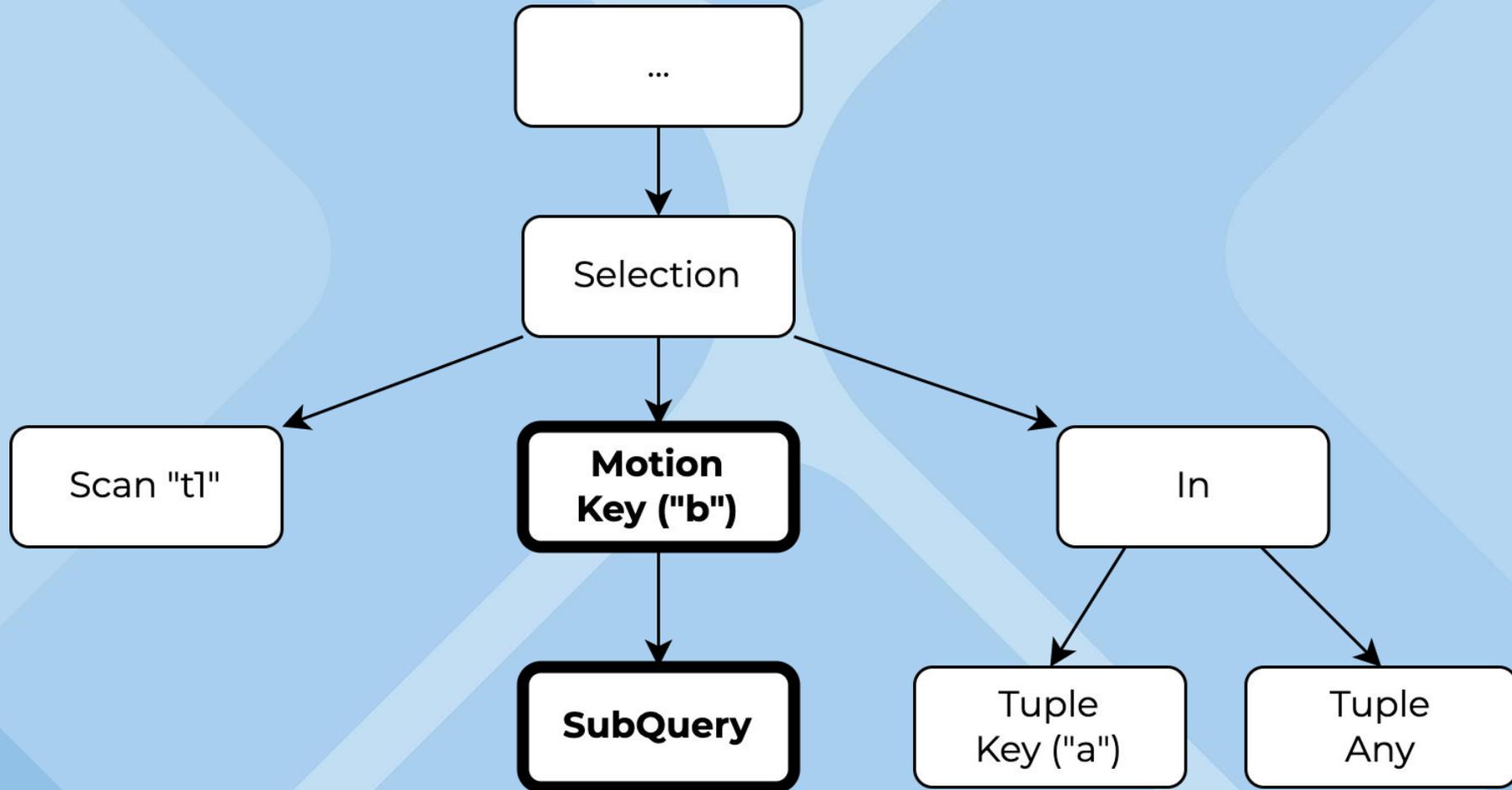
**SELECT a FROM t1
WHERE a IN (...)**





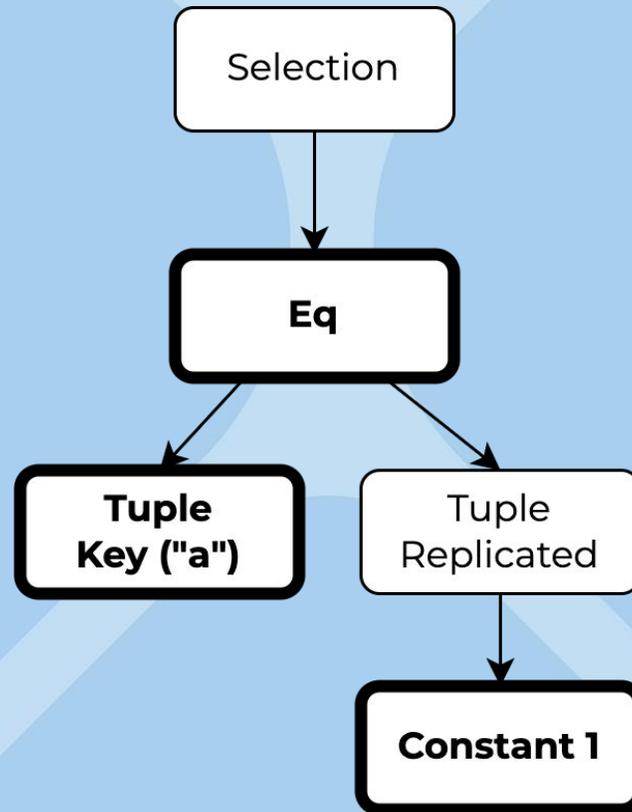
PICODATA

Motion



Discovery

**SELECT b FROM t2
WHERE a = 1**

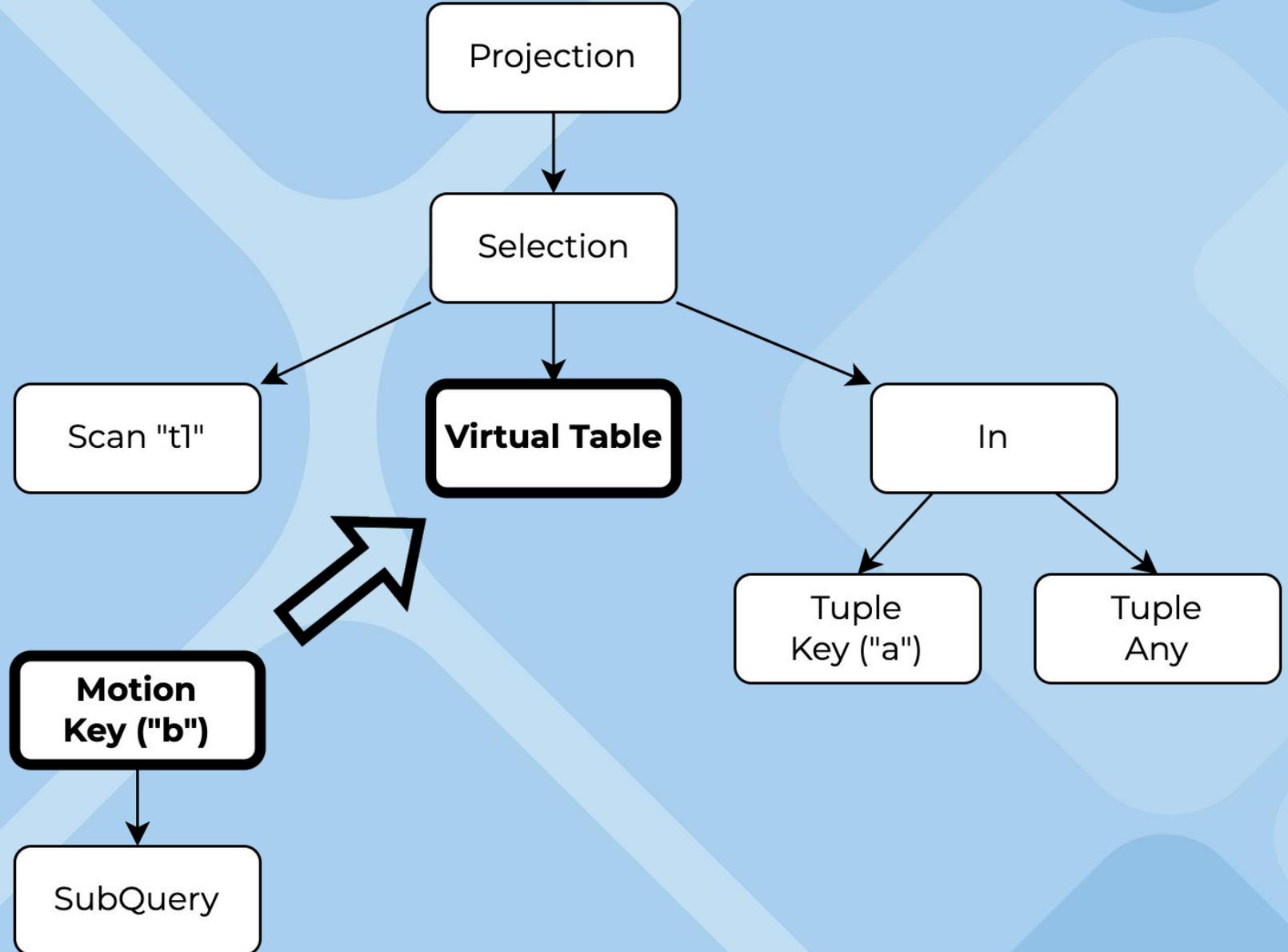




PICODATA

Linker

**SELECT a FROM t1
WHERE a IN
(VALUES (2), (3))**



Thank you!