

JOINS

MAP SIDE JOINS

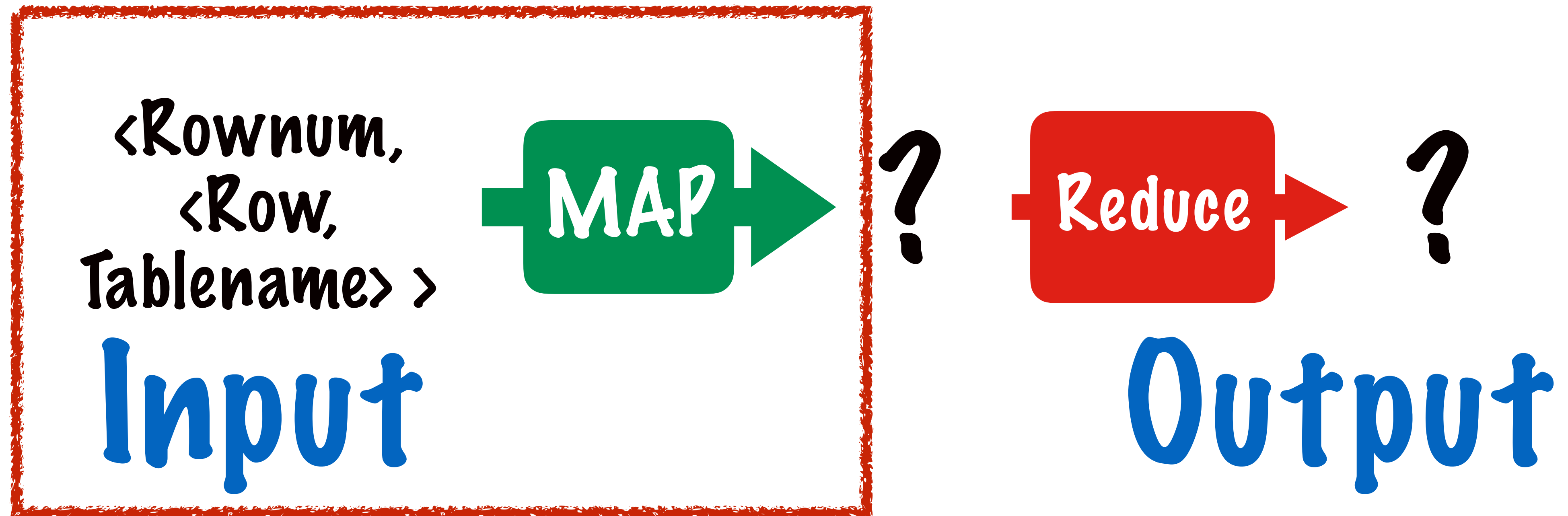
JOINS

MAP SIDE JOINS

A MAP SIDE JOIN IS THAT JOIN WHICH IS PERFORMED AS A **MAP ONLY** JOB

THE MAP SIDE JOIN DOESN'T
NEED A **REDUCER**

MAP SIDE JOINS



THE JOIN ONLY USES MAPPERS

MAP SIDE JOINS

THE JOIN ONLY USES MAPPERS

THERE ARE TWO CIRCUMSTANCES
WHICH RESULT IN MAP SIDE JOINS

1ST CONDITION - IF ALL BUT ONE
OF THE TABLES ARE SMALL

MAP SIDE JOINS

1ST CONDITION - IF ALL BUT ONE OF
THE TABLES ARE SMALL

THE SMALLER TABLES ARE
READ FROM HDFS AND STORED
IN AN IN-MEMORY HASH-TABLE

MAP SIDE JOINS

1ST CONDITION - IF ALL BUT ONE
OF THE TABLES ARE SMALL

THE SMALLER TABLES ARE
READ FROM HDFS AND STORED
IN AN IN-MEMORY HASH-TABLE

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER'S LOCAL DISK

MAP SIDE JOINS

1ST CONDITION - IF ALL BUT ONE
OF THE TABLES ARE SMALL

THE SMALLER TABLES ARE
READ FROM HDFS AND STORED
IN AN IN-MEMORY HASH-TABLE

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER'S LOCAL DISK

THIS IS ACCOMPLISHED USING THE HADOOP
DISTRIBUTED CACHE

MAP SIDE JOINS

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER'S LOCAL DISK

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

EXECUTE A JOIN

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```


MAP SIDE JOINS

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER'S LOCAL DISK

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

LET US ASSUME NAMES IS THE
SMALLER TABLE

MAP SIDE JOINS

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER'S LOCAL DISK

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

LET US ASSUME NAMES
IS THE **SMALLER** TABLE

TRADES TABLE IS
DIVIDED INTO CHUNKS
AND **DISTRIBUTED**
OVER THE MAPPERS

1ST CONDITION - IF ALL BUT ONE OF THE TABLES ARE SMALL

THIS IN-MEMORY HASH-TABLE IS COPIED
ONTO EACH MAPPER' LOCAL DISK

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

TRADES

TRADES TABLE IS DIVIDED INTO CHUNKS AND DISTRIBUTED OVER THE MAPPERS

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIES	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIES	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



TRADES TABLE IS DIVIDED INTO CHUNKS AND **DISTRIBUTED** OVER THE MAPPERS
THIS IN-MEMORY HASH-TABLE IS COPIED ONTO EACH MAPPER'
LOCAL DISK

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIES	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIES	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIES	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIES	OPEN	HIGH
NESTELIN	EQ	33	33.4



SYMBOL	SERIES	OPEN	HIGH
--------	--------	------	------

TRADES TABLE
THIS

NAMES IS OUR IN-MEMORY HASH
TABLE WHICH IS COPIED ONTO EACH
MAPPER LOCAL DISK

THE MAPPERS
'

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7



SYMBOL	SERIES	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIES	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIES	OPEN	HIGH
NESTELIN	EQ	33	33.4



SYMBOL	SERIES	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5

SYMBOL	SERIES	OPEN	HIGH
--------	--------	------	------

TRADES

TABLE IS BROKEN INTO CHUNKS

AND DISTRIBUTED OVER ALL THE MAPPERS

MAPPERS

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

MAP

SYMBOL	SERIES	OPEN	HIGH
RIL	EQ	23	23.5

MAP

SYMBOL	SERIES	OPEN	HIGH
NESTELIN	EQ	33	33.4
ESPN	EQ	44	47.8

MAP

SYMBOL	SERIES	OPEN	HIGH
TATA	EQ	102	109
SONY	EQ	111	112

ALL THE MAPPERS TOGETHER WILL GENERATE THE FINAL OUTPUT

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIES	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIES	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIES	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIES	OPEN	HIGH
NESTELIN	EQ	33	33.4



SYMBOL	SERIES	OPEN	HIGH
--------	--------	------	------

**ALL THE MAPPERS TOGETHER WILL
GENERATE THE FINAL OUTPUT**

**IN A MAP SIDE JOIN, THE SMALL
TABLE NEEDS TO BE READ **JUST ONCE****

ALL THE MAPPERS TOGETHER WILL
GENERATE THE FINAL OUTPUT

MAP-SIDE JOINS GREATLY **REDUCE**
THE OVERHEAD INCURRED IN THE
SORT/MERGE AND REDUCE STAGES

ALL THE MAPPERS TOGETHER WILL
GENERATE THE FINAL OUTPUT

MAP-SIDE JOIN WORKS ONLY
WHEN ONE TABLE IS SIGNIFICANTLY
SMALLER THAN THE OTHER TABLE

WHETHER A JOIN WILL BE A MAP-SIDE
JOIN **DEPENDS ON THE TYPE OF JOIN**

INNER JOIN

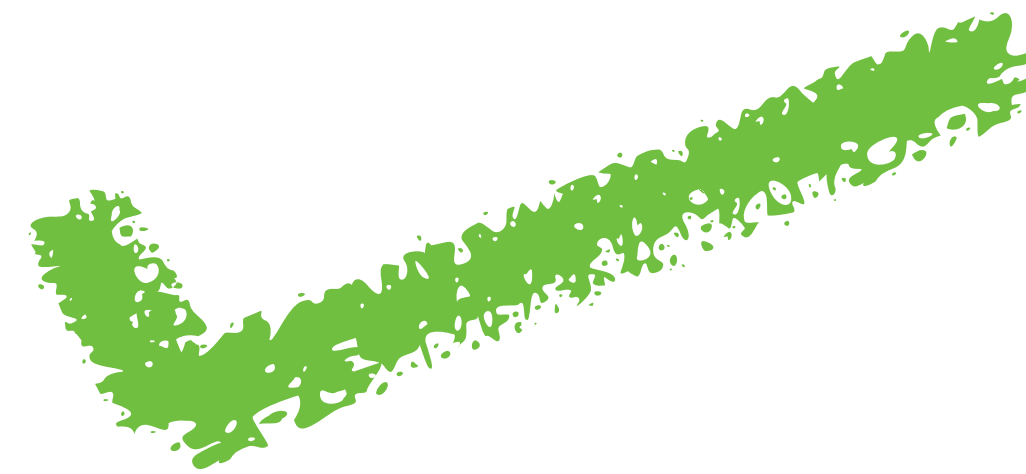
LEFT OUTER JOIN

RIGHT OUTER JOIN

FULL OUTER JOIN

WHETHER A JOIN WILL BE A MAP-SIDE
JOIN **DEPENDS ON THE TYPE OF JOIN**

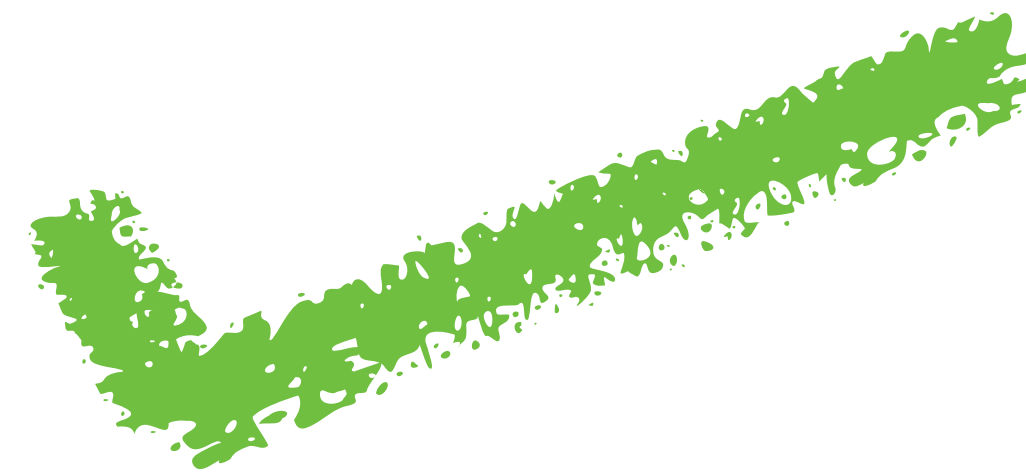
INNER JOIN



WE JUST SAW HOW INNER JOIN
WILL BE A MAP JOIN IF ONE
TABLE IS SMALLER

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

INNER JOIN



LEFT OUTER JOIN

RIGHT OUTER JOIN

FULL OUTER JOIN

LEFT OUTER JOIN

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES LEFT OUTER JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

NAMES IS THE SMALLER TABLE
AND **TRADES** IS THE BIGGER TABLE

LEFT OUTER JOIN

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES LEFT OUTER JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

NAMES IS THE SMALLER TABLE AND **TRADES** IS THE BIGGER TABLE

EVERY ROW OF THE LEFT TABLE (NAMES) SHOULD
BE PRESENT IN THE OUTPUT

WE HAVE ADDED FLIPKART
ROW WHICH WAS NOT
INITIALLY PRESENT

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

LEFT OUTER JOIN

EVERY ROW OF THE LEFT TABLE
SHOULD BE PRESENT IN THE OUTPUT

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

NAMES	
SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART
NAMES	
SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART
NAMES	
SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

TRADES

SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIES	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4



SYMBOL	SERIE	OPEN	HIGH
	S		

LEFT OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

ONE MAPPER HAS ONLY PARTIAL
INFORMATION OF THE RIGHT TABLE

NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART
NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART
NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART

TRADE				
SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4



SYMBOL	SERIE	OPEN	HIGH

LEFT OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

ONE MAPPER HAS ONLY PARTIAL
INFORMATION OF THE RIGHT TABLE

WE CAN'T TELL WHICH ROWS FROM
THE LEFT ARE **UNMATCHED!**

NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART
NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART
NAME	SYMBOL	NAME
	RIL	Reliance
	NESTLEIND	Nestle
	FLPK	FLIPKART

TRADE				
SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4




SYMBOL	SERIE	OPEN	HIGH

LEFT OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

ONE MAPPER HAS ONLY PARTIAL
INFORMATION OF THE RIGHT TABLE

WE CAN'T TELL WHICH ROWS FROM
THE LEFT ARE **UNMATCHED**!



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4

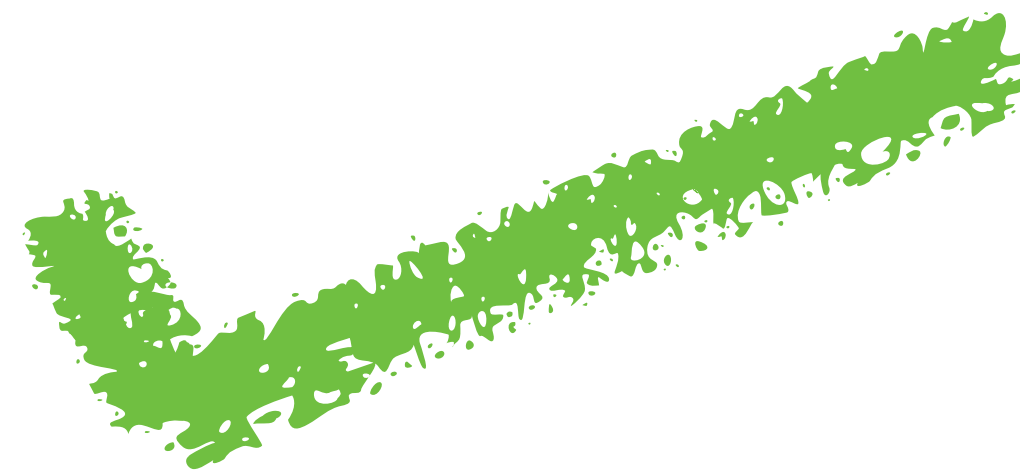


SYMBOL	SERIE	OPEN	HIGH

THE ROW WITH
FLIPKART IS UNMATCHED
BUT NO ONE MAPPER
HAS THAT INFORMATION!

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

INNER JOIN



LEFT OUTER JOIN

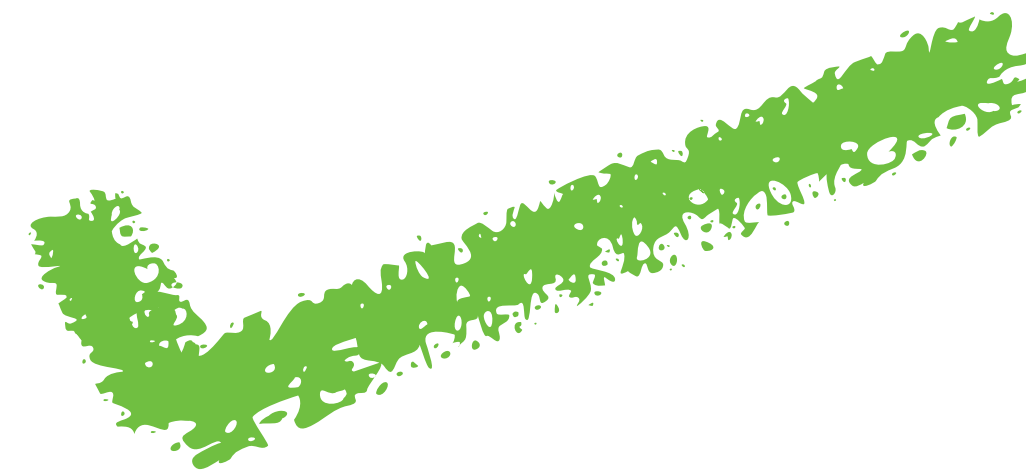


RIGHT OUTER JOIN

FULL OUTER JOIN

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

INNER JOIN



LEFT OUTER JOIN



RIGHT OUTER JOIN

FULL OUTER JOIN

RIGHT OUTER JOIN

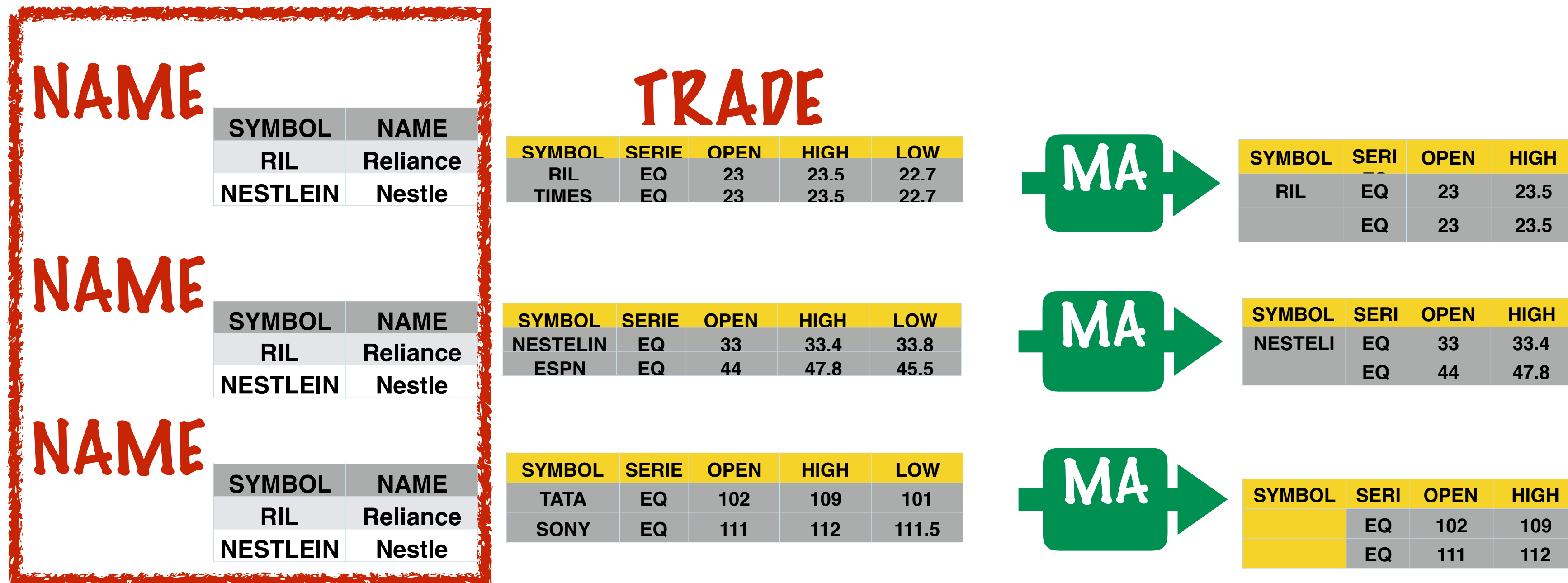
```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES RIGHT OUTER JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

NAMES IS THE SMALLER TABLE
AND TRADES IS THE BIGGER TABLE

RIGHT OUTER JOIN

EVERY ROW OF THE RIGHT TABLE IS
PRESENT IN THE OUTPUT

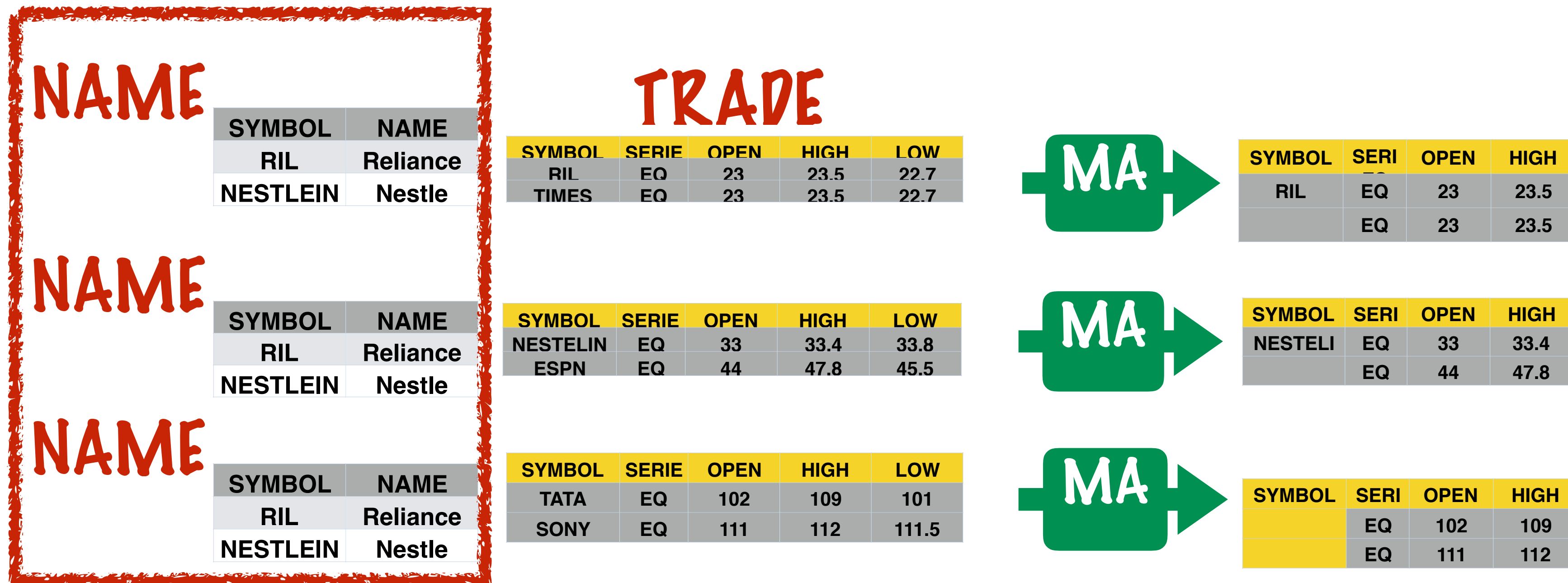
IF A ROW OF THE RIGHT TABLE DOES NOT MATCH
THE LEFT TABLE IT IS PADDED WITH NULLS



RIGHT OUTER JOIN

IF A ROW OF THE RIGHT TABLE DOES NOT MATCH
THE LEFT TABLE IT IS PADDED WITH NULLS

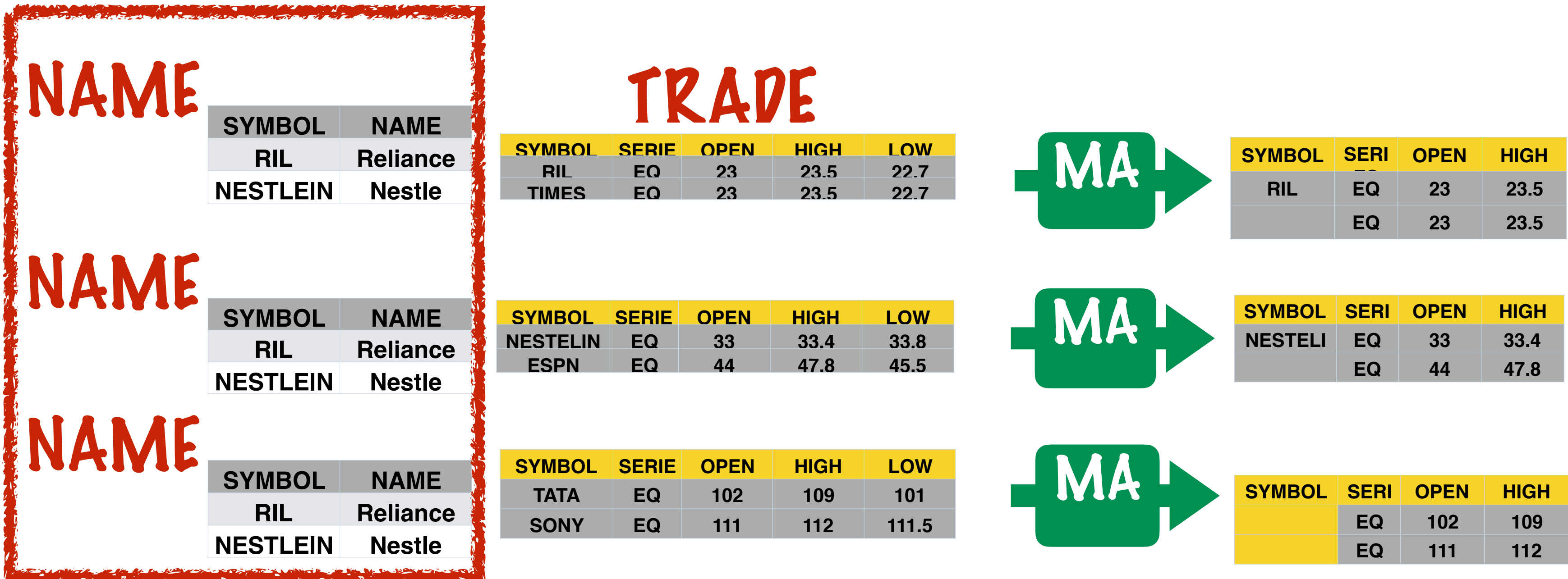
ONE MAPPER HAS ONLY PARTIAL INFORMATION
OF THE RIGHT TABLE **BUT THAT IS OKAY!**



RIGHT OUTER JOIN

ONE MAPPER HAS ONLY PARTIAL INFORMATION OF THE RIGHT TABLE **BUT THAT IS OKAY!**

EVERY MAPPER HAS COMPLETE INFORMATION OF THE LEFT TABLE



RIGHT OUTER JOIN

ONE MAPPER HAS ONLY PARTIAL INFORMATION OF THE RIGHT TABLE **BUT THAT IS OKAY!**

EVERY MAPPER HAS COMPLETE INFORMATION OF THE LEFT TABLE

EVERY ROW IN THE RIGHT TABLE CAN BE JOINED OR PADDED WITH NULLS

NAME

SYMBOL	NAME
RIL	Reliance
NESTLEIN	Nestle

NAME

SYMBOL	NAME
RIL	Reliance
NESTLEIN	Nestle

NAME

SYMBOL	NAME
RIL	Reliance
NESTLEIN	Nestle

TRADE

SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERI	OPEN	HIGH
RIL	EQ	23	23.5
	EQ	23	23.5



SYMBOL	SERI	OPEN	HIGH
NESTELI	EQ	33	33.4
	EQ	44	47.8



SYMBOL	SERI	OPEN	HIGH
	EQ	102	109
	EQ	111	112

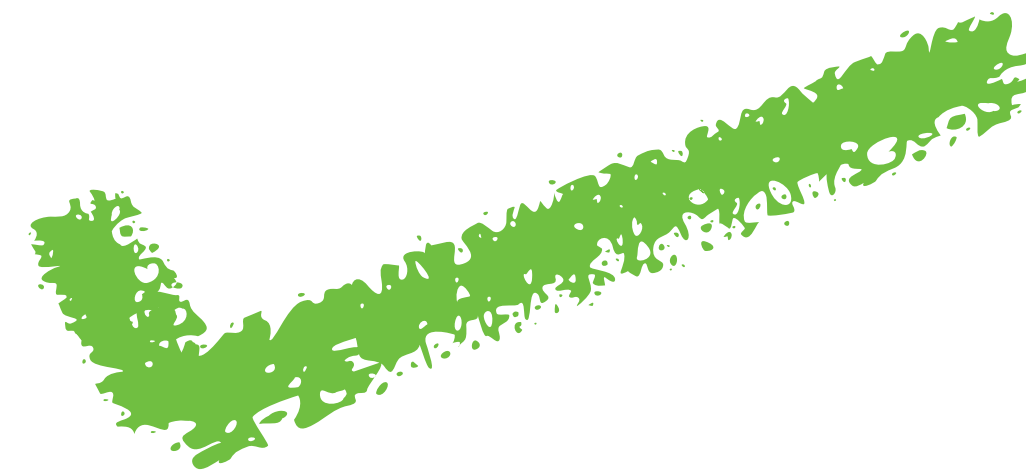
RIGHT OUTER JOIN

WHEN THE LEFT TABLE IS HELD IN
MEMORY I.E. IS THE SMALLER TABLE

RIGHT OUTER JOIN CAN
WORK AS A MAP JOIN!

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

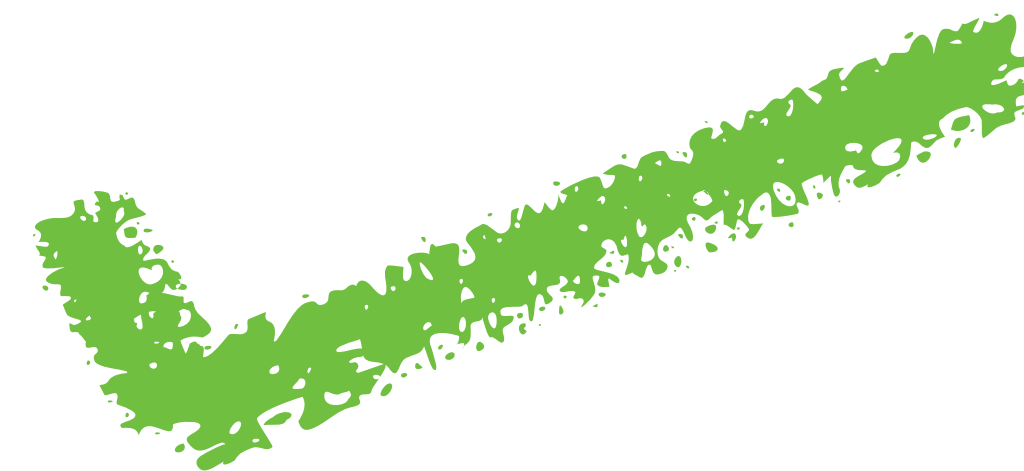
INNER JOIN



LEFT OUTER JOIN



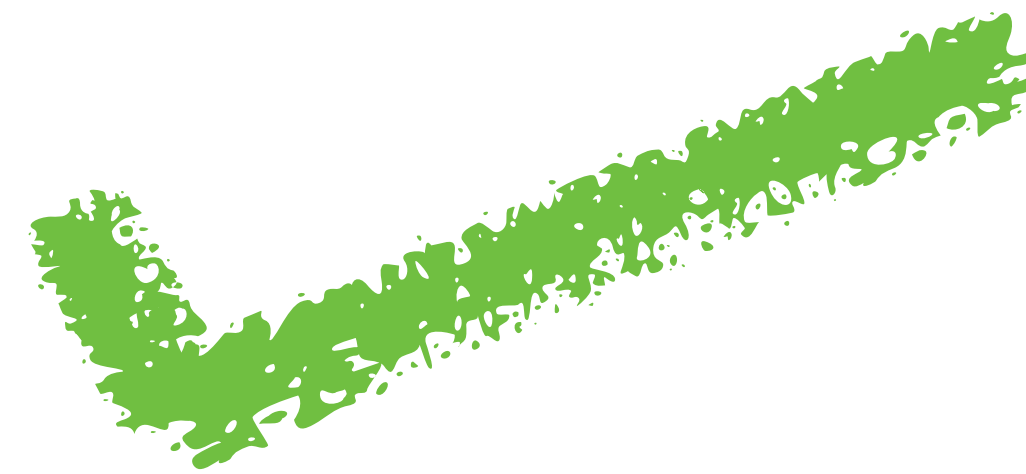
RIGHT OUTER JOIN



FULL OUTER JOIN

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

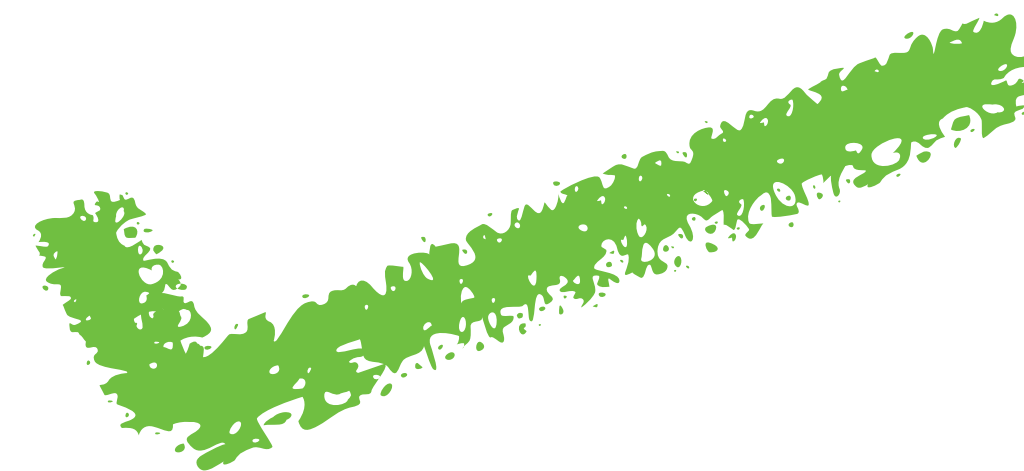
INNER JOIN



LEFT OUTER JOIN



RIGHT OUTER JOIN



FULL OUTER JOIN

FULL OUTER JOIN

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES FULL OUTER JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

NAMES IS THE SMALLER TABLE AND
TRADES IS THE BIGGER TABLE

FULL OUTER JOIN

```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES FULL OUTER JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

NAMES IS THE SMALLER TABLE AND **TRADES** IS THE BIGGER TABLE

EVERY ROW OF BOTH TABLES SHOULD
BE PRESENT IN THE OUTPUT

WE HAVE ADDED FLIPKART
ROW WHICH WAS NOT
INITIALLY PRESENT

NAMES	
SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

FULL OUTER JOIN

EVERY ROW OF THE LEFT TABLE
SHOULD BE PRESENT IN THE OUTPUT

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

TRADE

SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5
	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4
	EQ	44	47.8



SYMBOL	SERIE	OPEN	HIGH
	EQ	102	109
	EQ	111	112

FULL OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

EVERY MAPPER HAS ONLY PARTIAL
INFORMATION OF THE RIGHT TABLE

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

TRADE

SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5
	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4
	EQ	44	47.8



SYMBOL	SERIE	OPEN	HIGH
	EQ	102	109
	EQ	111	112

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5

FULL OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH THE RIGHT TABLE IT IS PADDED WITH NULLS

EVERY MAPPER HAS ONLY PARTIAL INFORMATION OF THE RIGHT TABLE

WE CAN'T TELL WHICH ROWS FROM THE LEFT ARE UNMATCHED!

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle
FLPK	FLIPKART

TRADE

SYMBOL	SERIE	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5
	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4
	EQ	44	47.8



SYMBOL	SERIE	OPEN	HIGH
	EQ	102	109
	EQ	111	112

SYMBOL	SERIE	OPEN	HIGH	LOW
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5

SYMBOL	SERIE	OPEN	HIGH	LOW
TATA	EQ	102	109	101
SONY	EQ	111	112	111.5

FULL OUTER JOIN

IF A ROW OF THE LEFT TABLE DOES NOT MATCH
THE RIGHT TABLE IT IS PADDED WITH NULLS

EVERY MAPPER HAS ONLY PARTIAL
INFORMATION OF THE RIGHT TABLE

WE CAN'T TELL WHICH ROWS FROM
THE LEFT ARE **UNMATCHED!**



SYMBOL	SERIE	OPEN	HIGH
RIL	EQ	23	23.5
	EQ	23	23.5



SYMBOL	SERIE	OPEN	HIGH
NESTELIN	EQ	33	33.4
	EQ	44	47.8

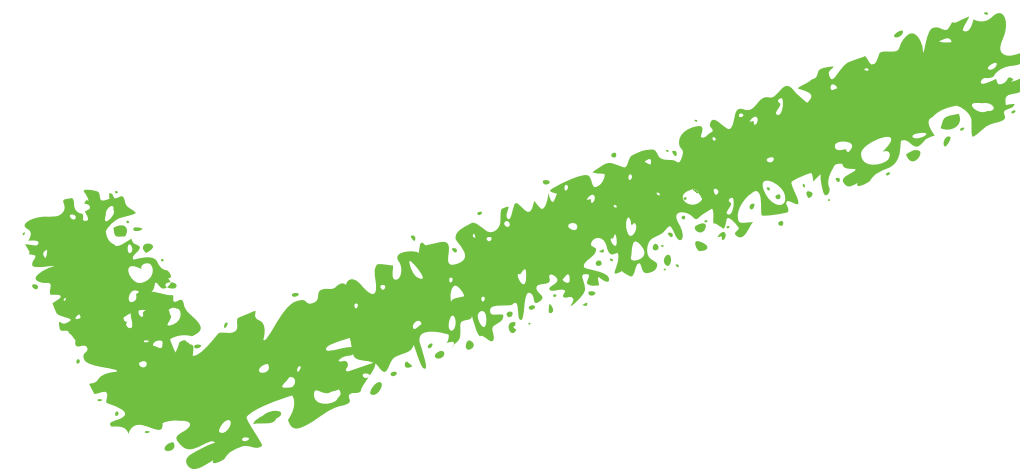


SYMBOL	SERIE	OPEN	HIGH
	EQ	102	109
	EQ	111	112

THE RESULT FROM
THE MAPPER
DOESN'T HAVE THE
FLIPKART ROW!

WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

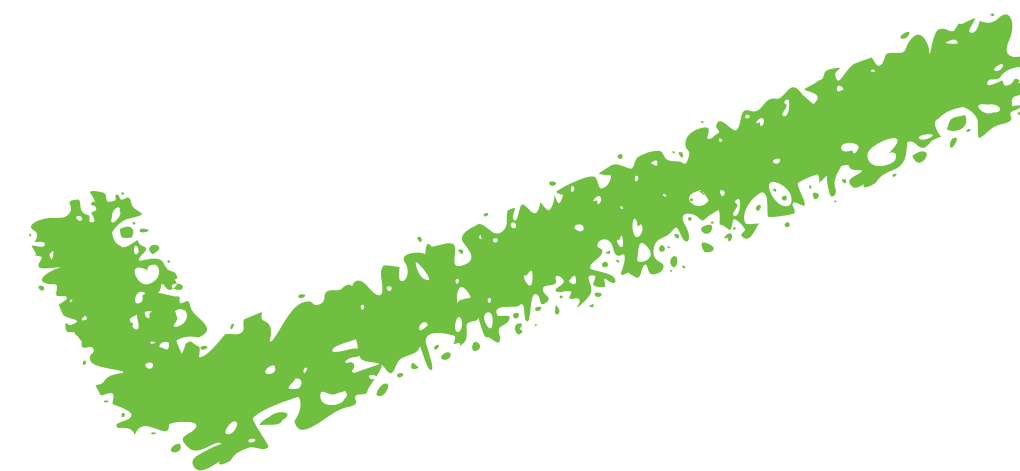
INNER JOIN



LEFT OUTER JOIN



RIGHT OUTER JOIN

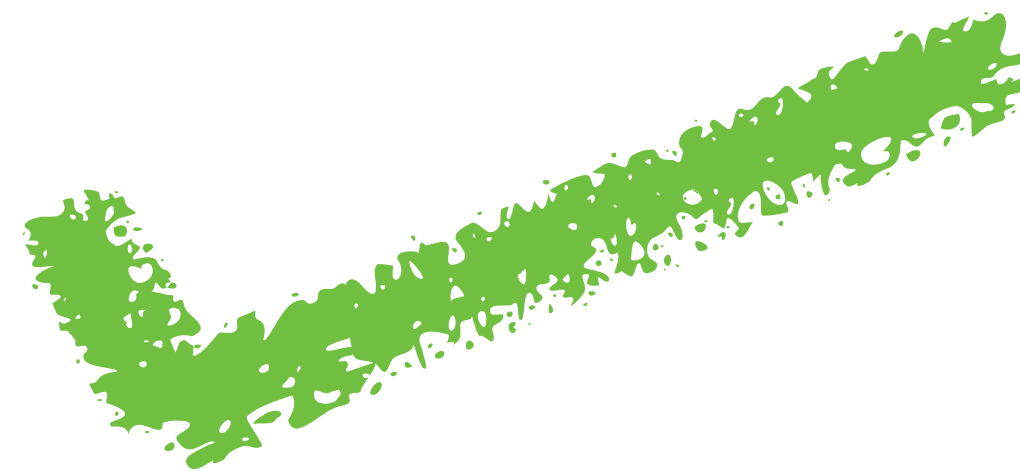


FULL OUTER JOIN



WHETHER A JOIN WILL BE A MAP-SIDE JOIN **DEPENDS ON THE TYPE OF JOIN**

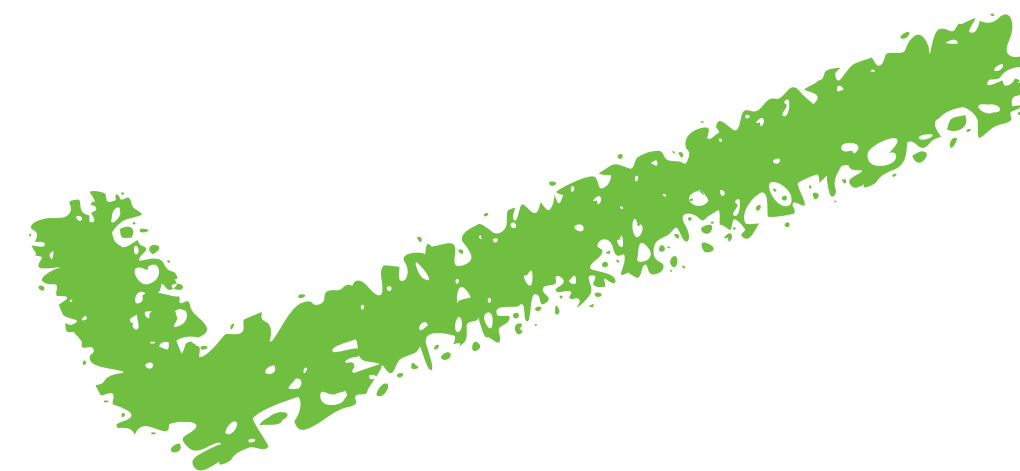
INNER JOIN



LEFT OUTER JOIN



RIGHT OUTER JOIN



FULL OUTER JOIN



SOMETHING TO NOTE

LEFT OUTER JOIN

WHEN THE **RIGHT** TABLE IS HELD IN
MEMORY I.E. IS THE SMALLER TABLE

LEFT OUTER JOIN CAN WORK
AS A MAP JOIN!

ONE CAN EXPLICITLY CALL OUT THAT
THE JOIN SHOULD BE A **MAP JOIN**

```
/*+ MAPJOIN (TRADES) */
```

```
SELECT /*+ MAPJOIN (TRADES) */  
NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

MAP SIDE JOINS

THE JOIN ONLY USES MAPPERS

THERE ARE TWO CIRCUMSTANCES IN WHICH MAP SIDE JOINS CAN HAPPEN

2ND CONDITION - IF TABLES ARE BUCKETED ON THE JOIN COLUMNS AND THE NUMBER OF BUCKETS IN ONE TABLE IS A MULTIPLE OF THE NUMBER OF BUCKETS IN THE SECOND TABLE

MAP SIDE JOINS

2ND CONDITION - IF TABLES ARE BUCKETED ON THE JOIN COLUMNS AND THE NUMBER OF BUCKETS IN ONE TABLE IS A MULTIPLE OF THE NUMBER OF BUCKETS IN THE SECOND TABLE

LET US ASSUME THERE ARE TWO TABLES

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

MAP SIDE JOINS

ND
2 CONDITION - IF TABLES ARE BUCKETED ON THE JOIN COLUMNS AND THE NUMBER OF BUCKETS IN ONE TABLE IS A MULTIPLE OF THE NUMBER OF BUCKETS IN THE SECOND TABLE

LET US ASSUME THERE ARE TWO TABLES

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

MAP SIDE JOINS

ND

2 CONDITION - IF TABLES ARE BUCKETED ON THE JOIN COLUMNS AND THE NUMBER OF BUCKETS IN ONE TABLE IS A MULTIPLE OF THE NUMBER OF BUCKETS IN THE SECOND TABLE

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

BOTH OF THEM ARE BUCKETED ON SAME COLUMN - SYMBOL

THE JOIN IS ALSO ON 'SYMBOL'

```
SELECT NAMES . SYMBOL , SERIES , HIGH
FROM NAMES JOIN TRADES
ON (NAMES . SYMBOL = TRADES . SYMBOL) ;
```

MAP SIDE JOINS

ND

2 CONDITION - IF TABLES ARE BUCKETED ON THE JOIN COLUMNS AND THE NUMBER OF BUCKETS IN ONE TABLE IS A MULTIPLE OF THE NUMBER OF BUCKETS IN THE SECOND TABLE

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
--------	--------	------	------	-----

**BOTH OF THEM ARE BUCKETED ON
SAME COLUMN - SYMBOL**

THE JOIN IS ALSO ON 'SYMBOL'

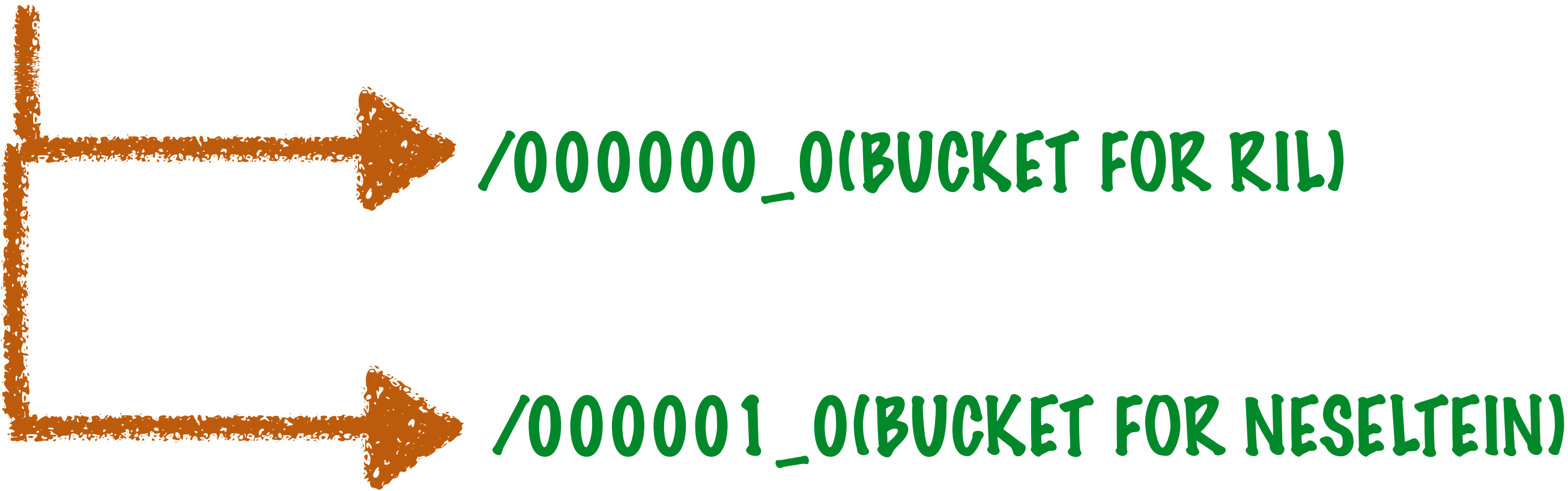
```
SELECT NAMES.SYMBOL, SERIES, HIGH  
FROM NAMES JOIN TRADES  
ON (NAMES.SYMBOL = TRADES.SYMBOL) ;
```

BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

/USER/HIVE/WAREHOUSE
/NAMES

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle



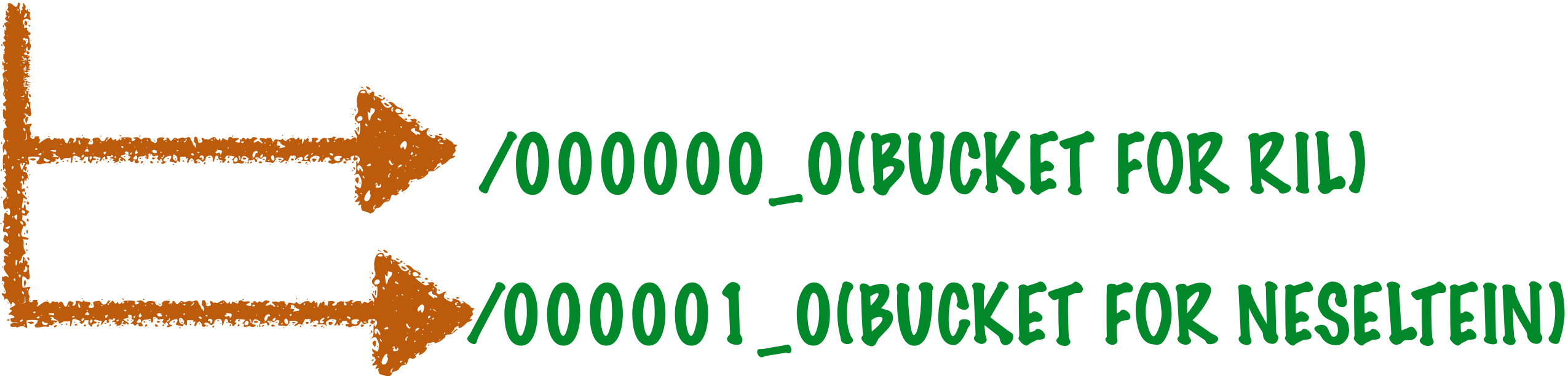
BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

NAMES

SYMBOL	NAME
RIL	Reliance
NESTLEIND	Nestle

/USER/HIVE/WAREHOUSE

/NAMES

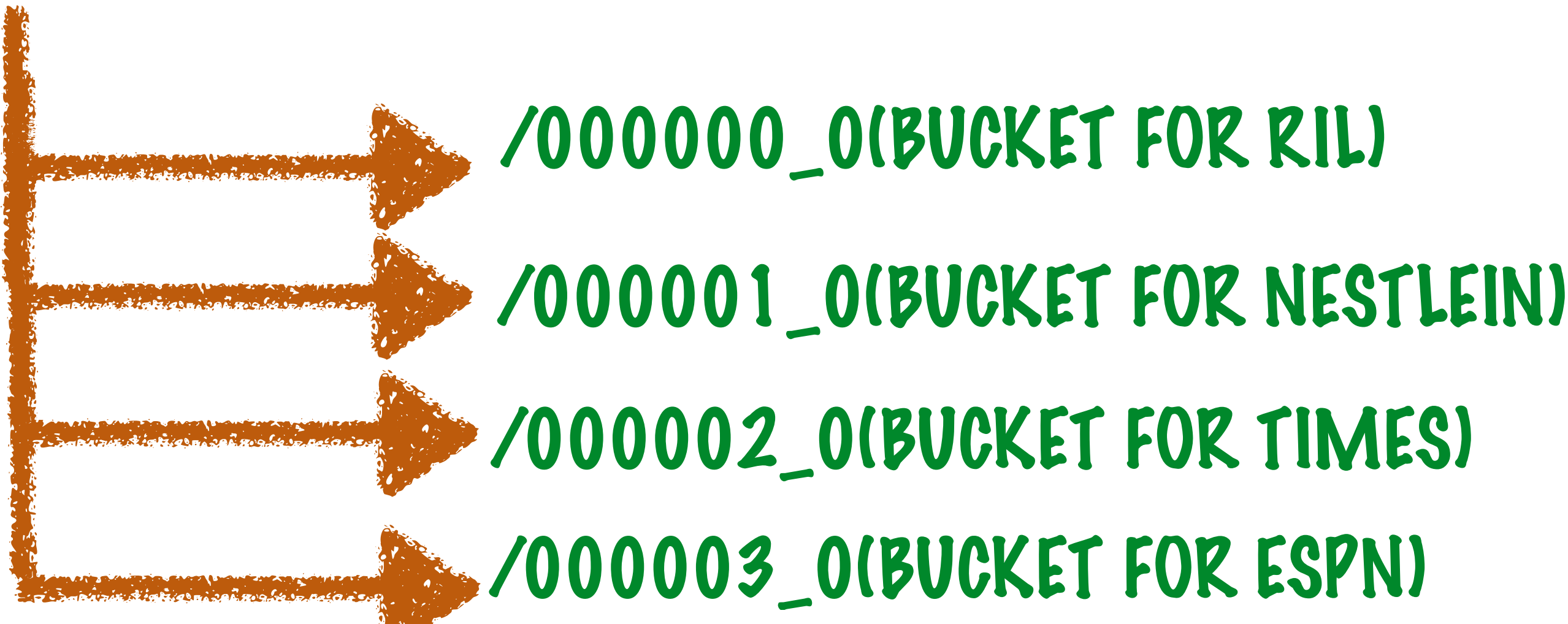


/USER/HIVE/WAREHOUSE

/TRADES

TRADES

SYMBOL	SERIES	OPEN	HIGH	LOW
RIL	EQ	23	23.5	22.7
TIMES	EQ	23	23.5	22.7
NESTELIN	EQ	33	33.4	33.8
ESPN	EQ	44	47.8	45.5



BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

/USER/HIVE/WAREHOUSE

/NAMES

INSTEAD OF
FETCHING THE
ENTIRE



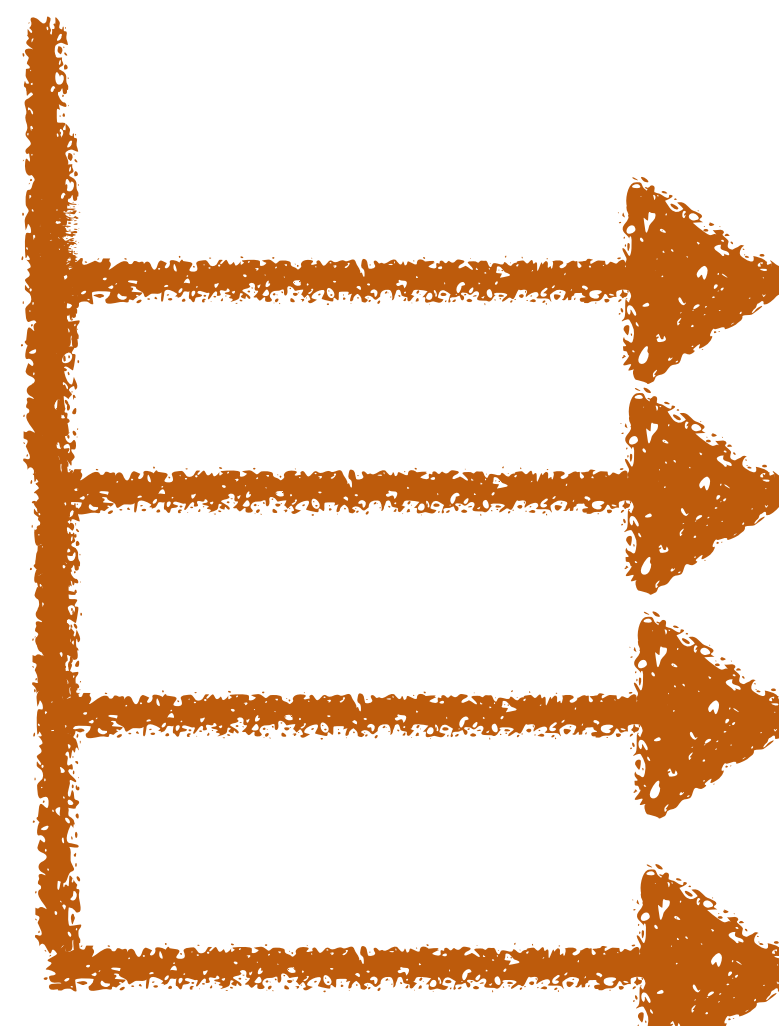
/000000_0(BUCKET FOR RIL)

/000001_0(BUCKET FOR NESELTEIN)

"TRADES" TABLE **/USER/HIVE/WAREHOUSE**

ONLY THE
REQUIRED
BUCKETS ARE
FETCHED

/TRADES



/000000_0(BUCKET FOR RIL)

/000001_0(BUCKET FOR NESTLEIN)

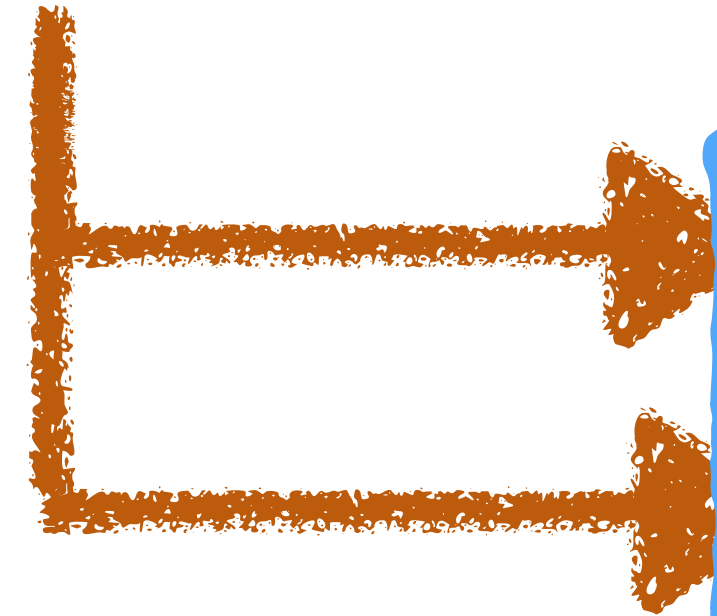
/000002_0(BUCKET FOR TIMES)

/000003_0(BUCKET FOR ESPN)

BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

/USER/HIVE/WAREHOUSE

/NAMES

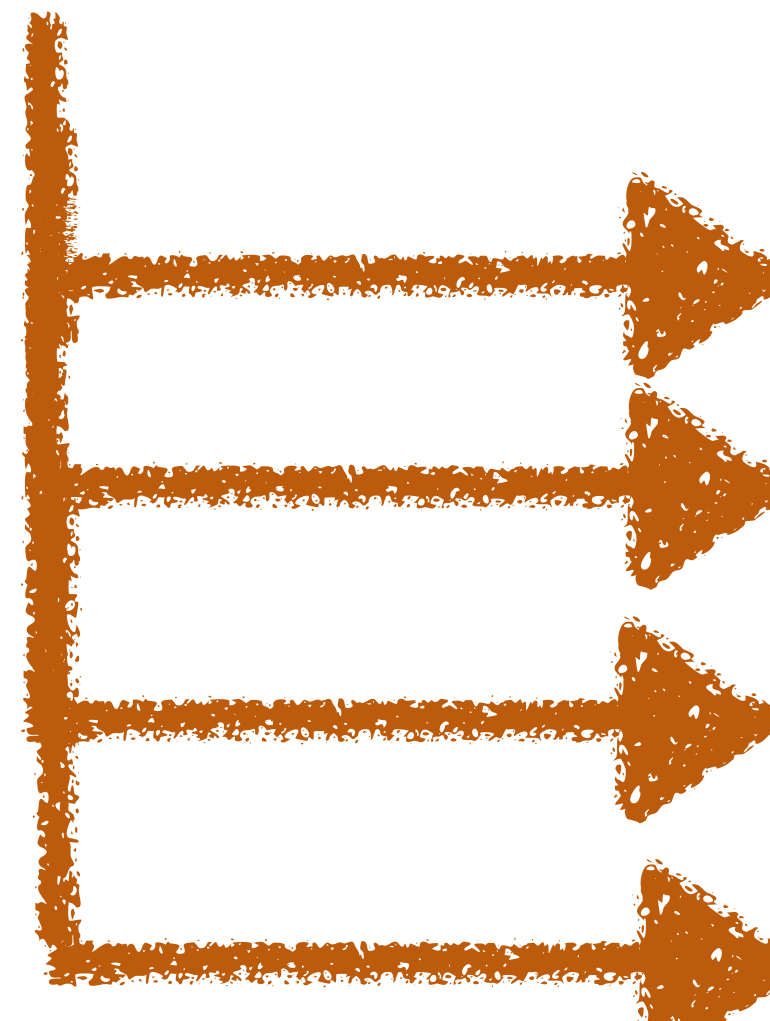


/000000_0(BUCKET FOR RIL)

/000001_0(BUCKET FOR NESELTEIN)

/USER/HIVE/WAREHOUSE

/TRADES



/000000_0(BUCKET FOR RIL)

/000001_0(BUCKET FOR NESTLEIN)

/000002_0(BUCKET FOR TIMES)

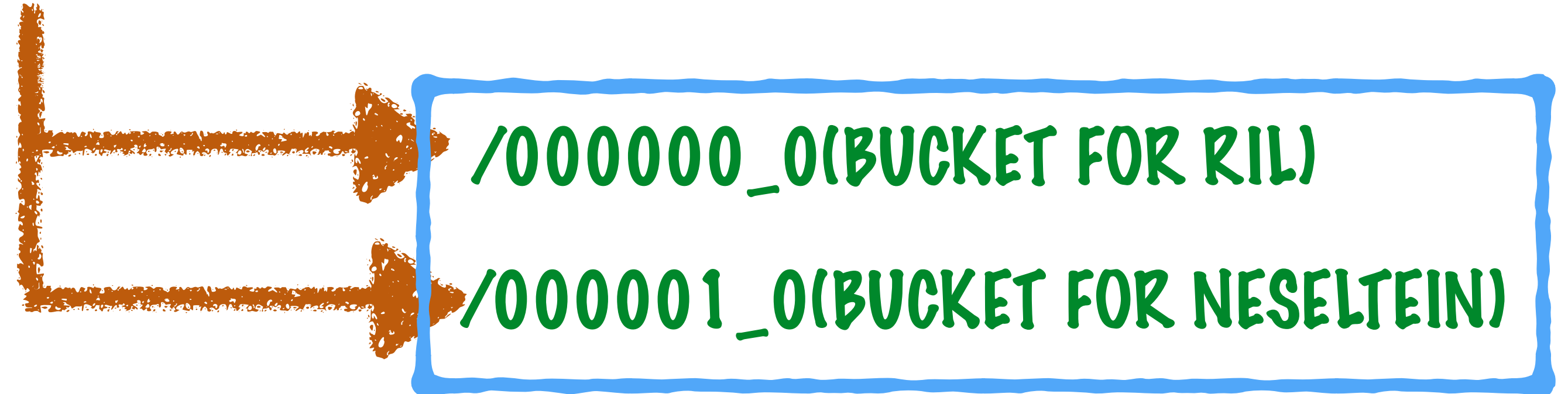
/000003_0(BUCKET FOR ESPN)

THESE ARE
THE ONLY
PARTS OF
THE TABLE
NEEDED

BOTH OF THEM ARE BUCKETED ON SAME COLUMN - **SYMBOL**

/USER/HIVE/WAREHOUSE

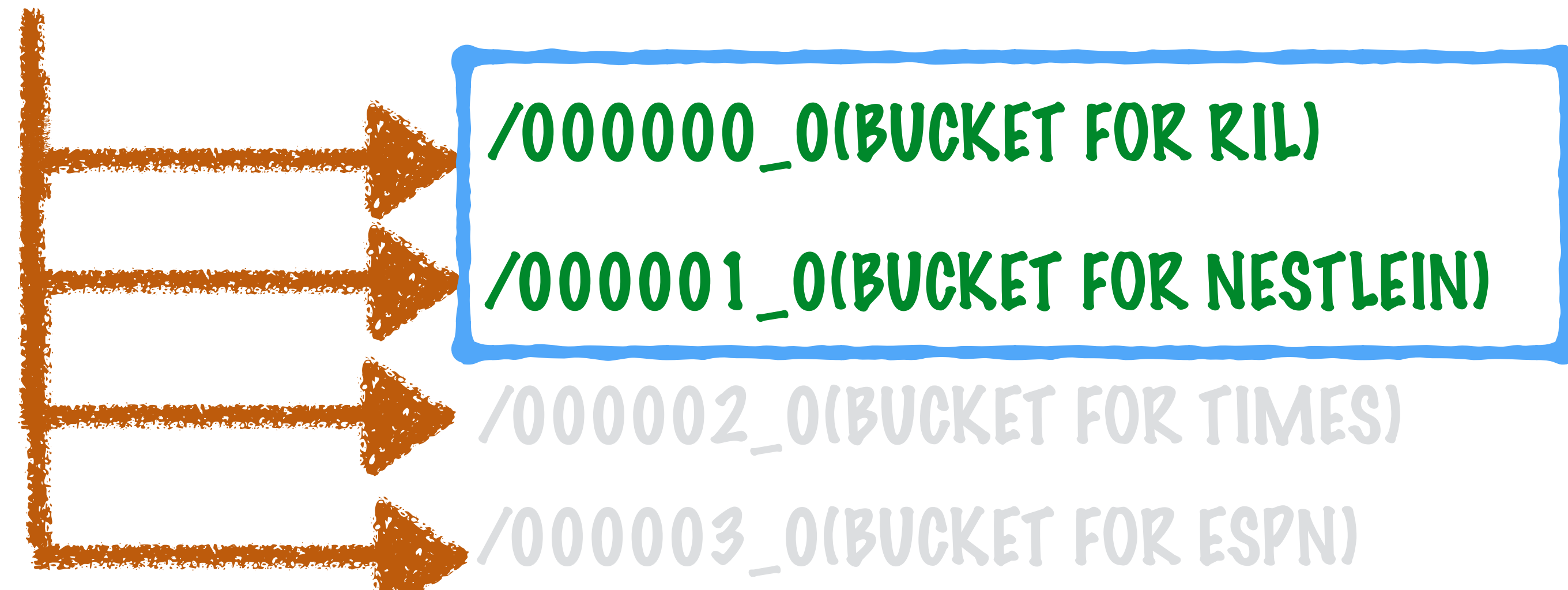
/NAMES



**BUCKET
MAP JOIN**

/USER/HIVE/WAREHOUSE

/TRADES



BUCKET MAP JOIN

BUCKET MAP JOINS ARE NOT ENABLED BY DEFAULT

```
set hive.optimize.bucketmapjoin = true;
```

BUCKET MAP JOIN

IF THE TABLES

ARE BUCKETED USING THE JOIN COLUMN

ARE SORTED USING THE JOIN COLUMN

AND

HAVE THE SAME NUMBER OF BUCKETS

BUCKET MAP JOIN

IF THE TABLES

ARE BUCKETED USING THE JOIN COLUMN

ARE SORTED USING THE JOIN COLUMN

AND

HAVE THE SAME NUMBER OF BUCKETS

THEN THE MAP-JOIN CAN BE MUCH
MORE EFFICIENT

BUCKET MAP JOIN

IF THE TABLES

ARE BUCKETED USING THE JOIN COLUMN

ARE SORTED USING THE JOIN COLUMN

AND

HAVE THE SAME NUMBER OF BUCKETS

MERGE-SORT

THE ROWS CAN BE JOINED USING
MERGE-SORT

SORT-MERGE JOINS CAN BE
USED ONLY FOR EQUIJOINS

THE ROWS CAN BE JOINED USING MERGE-SORT

SORT-MERGE IS NOT PERFORMED BY DEFAULT

```
set hive.input.format =  
org.apache.hadoop.hive ql.io.BucketizedHiveInputFormat;  
  
set hive.optimize.bucketmapjoin = true;  
  
set hive.optimize.bucketmapjoin.sortedmerge = true;
```

**IN SOME SCENARIOS, MAPJOINS ARE NOT
SUPPORTED**

UNION FOLLOWED BY A MAPJOIN

LATERAL VIEW FOLLOWED BY A MAPJOIN

WE WILL COVER LATERAL VIEW LATER

MAPJOIN FOLLOWED BY UNION

MAPJOIN FOLLOWED BY JOIN

MAPJOIN FOLLOWED BY MAPJOIN