



# LOOKING FOR DIFFERENCES IN YOUR DATA

TATIANA KRUPENYA

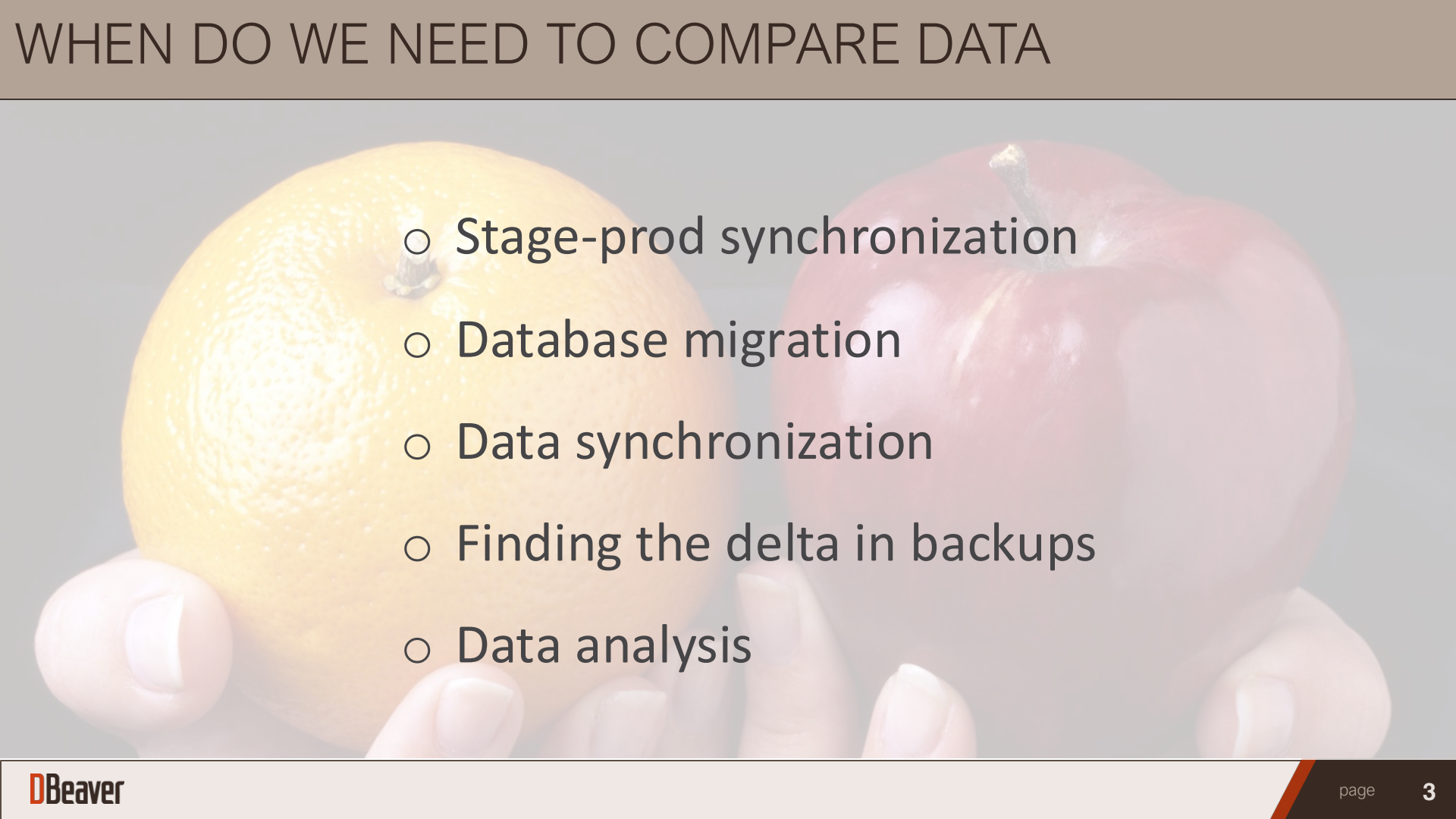
September 30, 2024

# NICE TO MEET YOU

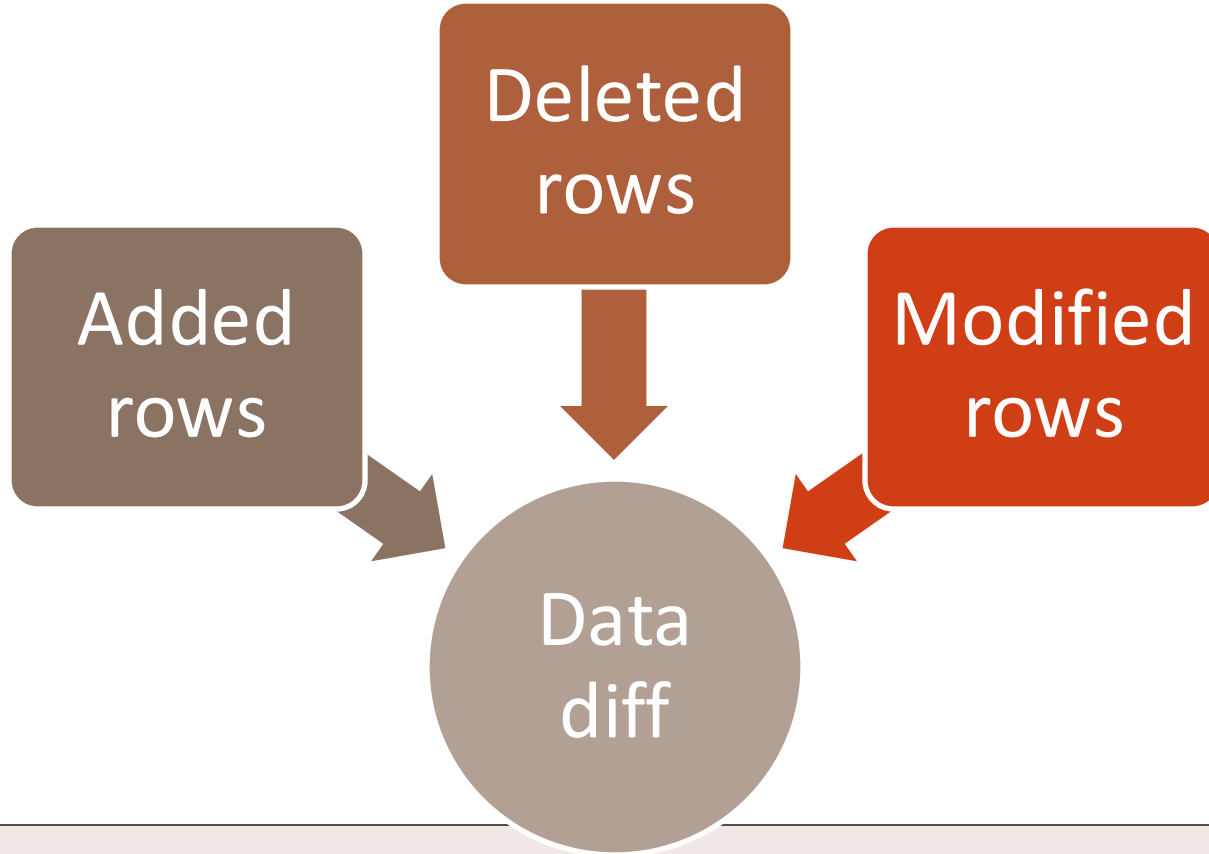
- I am the CEO of DBeaver
- DBeaver is a UI tool for database management and administration
- DBeaver was born as an open-source project in 2011
- The number of DBeaver users has grown to more than 8 million
- DBeaver is used in every country in the world



# WHEN DO WE NEED TO COMPARE DATA

- 
- A hand is shown holding two fruits: a bright yellow lemon on the left and a shiny red apple on the right. The background is a soft, out-of-focus grey. The list of items is overlaid on the right side of the image.
- Stage-prod synchronization
  - Database migration
  - Data synchronization
  - Finding the delta in backups
  - Data analysis

# WHAT ARE THE DIFFERENCES



# HOW TO COMPARE: BY EYES

```
> select * from employees e
```

employee_id	full_name	manager_id
1	Michael North	
2	Megan Berry	1
3	Sarah Berry	1
4	Zoe Black	1
5	Tim James	1
6	Bella Tucker	2
7	Ryan Metcalfe	2
8	Max Mills	2
9	Benjamin Glover	2
10	Carolyn Henderson	3
11	Nicola Kelly	3
12	Alexandra Climo	3
13	Dominic King	3
14	Leonard Gray	4
15	Eric Rampling	4
16	Piers Paige	7
17	Ryan Henderson	7
18	Frank Tucker	8
19	Nathan Ferguson	8
20	Kevin Rampling	8

20 row(s) fetched.



```
> select * from employees e
```

employee_id	full_name	manager_id
1	Michael North	
2	Megan Berry	1
3	Sarah Berry	1
4	Zoe Black	1
5	Tim James	1
6	Bella Tucker	2
7	Ryan Metcalfe	2
8	Max Mills	2
9	Benjamin Glover	2
10	Carolyn Henderson	3
11	Nicola Kelly	3
12	Alexandra Climo	3
13	Dominic King	3
14	Leonard Gray	4
15	Eric Rampling	4
16	Piers Paige	7
17	Ryan Henderson	7
18	Frank Tucker	8
19	Nathan Ferguson	8
20	Kevin Rampling	8
21	John Smith	5

21 row(s) fetched.

# HOW TO COMPARE: BY EYES

```
> select * from employees e
```

employee_id	full_name	manager_id
1	Michael North	
2	Megan Berry	1
3	Sarah Berry	1
4	Zoe Black	1
5	Tim James	1
6	Bella Tucker	2
7	Ryan Metcalfe	2
8	Max Mills	2
9	Benjamin Glover	2
10	Carolyn Henderson	3
11	Nicola Kelly	3
12	Alexandra Climo	3
13	Dominic King	3
14	Leonard Gray	4
15	Eric Rampling	4
16	Piers Paige	7
17	Ryan Henderson	7
18	Frank Tucker	8
19	Nathan Ferguson	8
20	Kevin Rampling	8

20 row(s) fetched.



```
> select * from employees e
```

employee_id	full_name	manager_id
1	Michael North	
2	Megan Berry	1
3	Sarah Berry	1
4	Zoe Black	1
5	Tim James	1
6	Bella Tucker	2
7	Ryan Metcalfe	2
8	Max Mills	2
9	Benjamin Glover	2
10	Carolyn Henderson	3
11	Nicola Kelly	3
12	Alexandra Climo	3
13	Dominic King	3
14	Leonard Gray	4
15	Eric Rampling	4
16	Piers Paige	7
17	Ryan Henderson	7
18	Frank Tucker	8
19	Nathan Ferguson	8
20	Kevin Rampling	8
21	John Smith	5

21 row(s) fetched.

- ✓ Simple queries
- ✓ Not require additional tools
- ✓ Not require knowledge
- ✗ For a small amount of data
- ✗ Diff is only in your head
- ✗ Better to have meaningful data

# HOW TO COMPARE: IN EXCEL

The image shows two Excel spreadsheets side-by-side, illustrating a comparison of fruit prices. The left spreadsheet (Book1 - 1) has a pink background for the price column, and the right spreadsheet (Book1 - 2) has a yellow background. An 'Edit Formatting Rule' dialog box is open on the right, showing a formula '=B2<>June!B2' and a 'Yellow Fill with Dark Yellow Text' format.

	A	B	C
1	Product	Price	
2	Apple	\$ 2.00	
3	Lemon	\$ 3.00	
4	Mango	\$ 4.00	
5	Lime	\$ 1.00	
6	Orange	\$ 2.00	
7	Plum	\$ 2.00	
8	Banana	\$ 0.50	
9	Pinapple	\$ 4.00	
10			

	A	B	C
1	Product	Price	
2	Apple	\$ 1.00	
3	Lemon	\$ 2.00	
4	Mango	\$ 4.00	
5	Lime	\$ 1.00	
6	Orange	\$ 2.00	
7	Plum	\$ 1.00	
8	Banana	\$ 0.50	
9	Pinapple	\$ 5.00	
10			

**Edit Formatting Rule**

Style: Classic

Use a formula to determine which cells to format

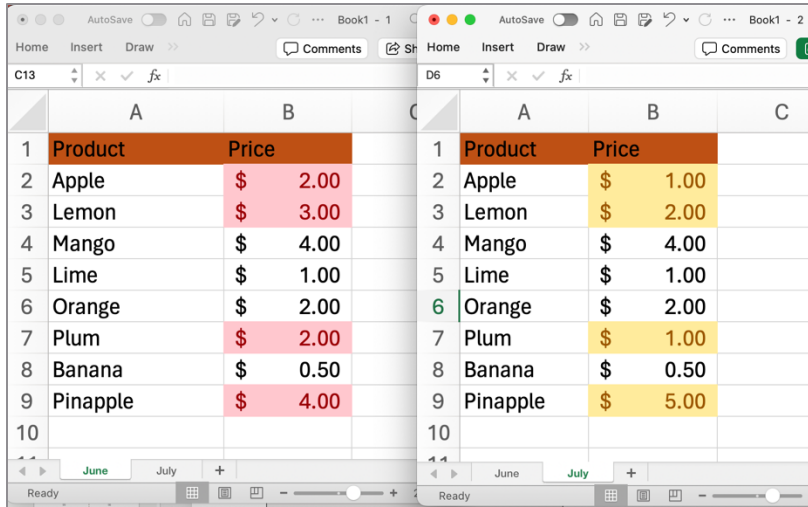
=B2<>June!B2

Format with: Yellow Fill with Dark Yellow Text

AaBbCcYyZz

Cancel OK

# HOW TO COMPARE: IN EXCEL



	A	B	C
1	Product	Price	
2	Apple	\$ 2.00	
3	Lemon	\$ 3.00	
4	Mango	\$ 4.00	
5	Lime	\$ 1.00	
6	Orange	\$ 2.00	
7	Plum	\$ 2.00	
8	Banana	\$ 0.50	
9	Pinapple	\$ 4.00	
10			

	A	B	C
1	Product	Price	
2	Apple	\$ 1.00	
3	Lemon	\$ 2.00	
4	Mango	\$ 4.00	
5	Lime	\$ 1.00	
6	Orange	\$ 2.00	
7	Plum	\$ 1.00	
8	Banana	\$ 0.50	
9	Pinapple	\$ 5.00	
10			

- ✓ Simple queries
- ✓ No manual work
- ✓ The diff is visible
  
- ✗ Need to export data first
- ✗ Need Excel or similar tools
- ✗ Need to know Excel magic



# HOW TO COMPARE: WITH SQL

MODIFIED

```
select lt.id, lt.uid lt_uid, lt."name" lt_name, rt.uid rt_uid, rt."name"  
rt_name from table_diff_1 lt  
join table_diff_2 rt on rt.id = lt.id and (rt.uid <> lt.uid or rt."name" <>  
lt."name");
```

DELETED

```
select lt.* from table_diff_1 lt  
where not exists (select 1 from table_diff_2 rt where rt.id = lt.id );
```

ADDED

```
select rt.* from table_diff_2 rt  
where not exists (select 1 from table_diff_1 lt where rt.id = lt.id );
```

# HOW TO COMPARE: WITH SQL

```
select lt.id, lt.uid lt_uid, lt."name" lt_name,  
rt.uid rt_uid, rt."name" rt_name from  
table_diff_1 lt  
join table_diff_2 rt on rt.id = lt.id and  
(rt.uid <> lt.uid or rt."name" <> lt."name");
```

```
select lt.* from table_diff_1 lt  
where not exists (select 1 from table_diff_2 rt  
where rt.id = lt.id );
```

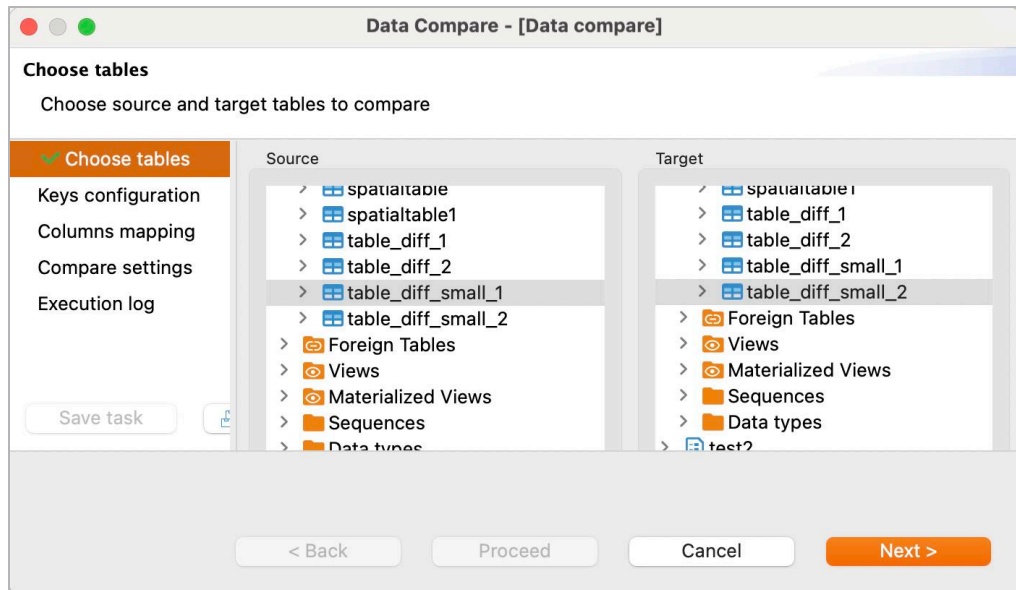
```
select rt.* from table_diff_2 rt  
where not exists (select 1 from table_diff_1 lt  
where rt.id = lt.id );
```

- ✓ No manual work
- ✓ The result is diff tables
- ✓ The nature of the data is not important
- ✗ Requires SQL knowledge
- ✗ Takes database resources
- ✗ May require FDW
- ✗ Not applicable to other databases

# HOW TO COMPARE: IN APP

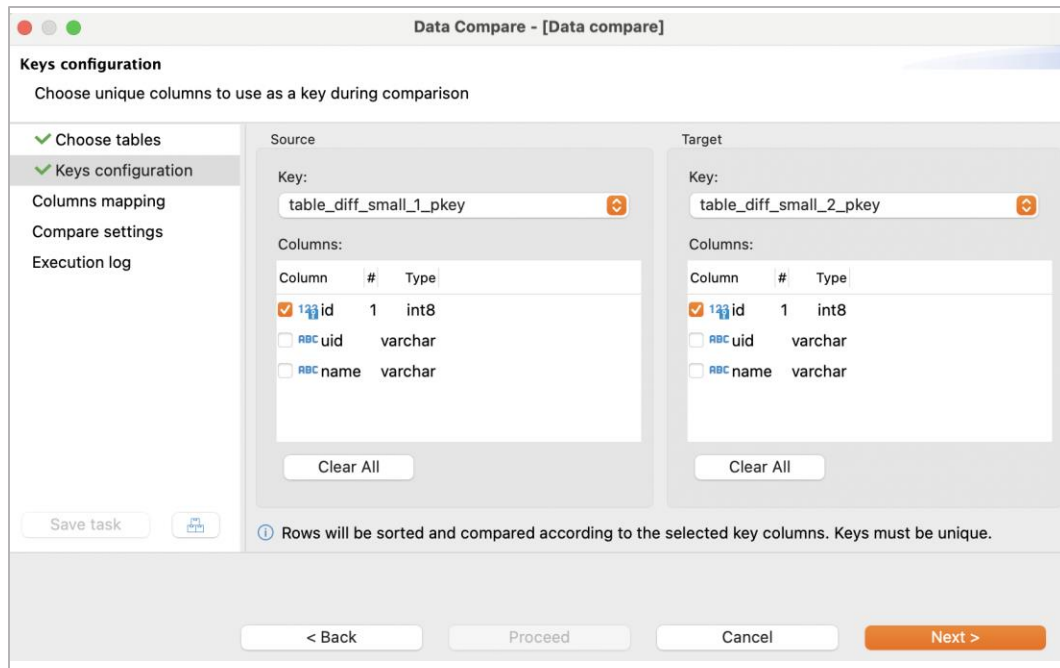
Source - test.table_diff_small_1 (PostgreSQL - gis)				Target - test.table_diff_small_2 (PostgreSQL - gis)				
	<small>123</small> id	ABC uid	ABC name		<small>123</small> id	ABC uid	ABC name	
1	0	new row	new row in left tab		1	35	valueX	changed value X
2	35	value35	load tests 35		2	135	valueZ	changed value Z
3	135	value135	load tests 135		3	1,001	missing row	missing row in left

# HOW TO COMPARE: IN APP



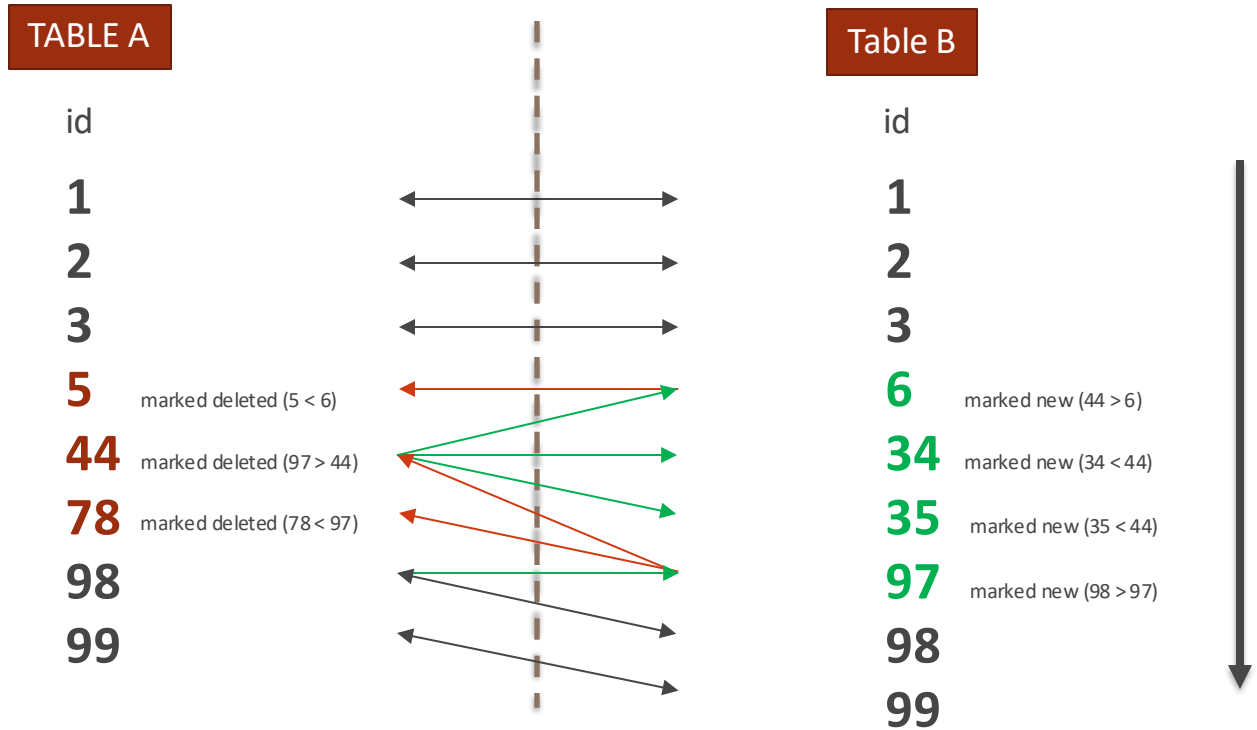
```
select * from table_diff_small_1 tds;  
select * from table_diff_small_2 tds;
```

# HOW TO COMPARE: IN APP

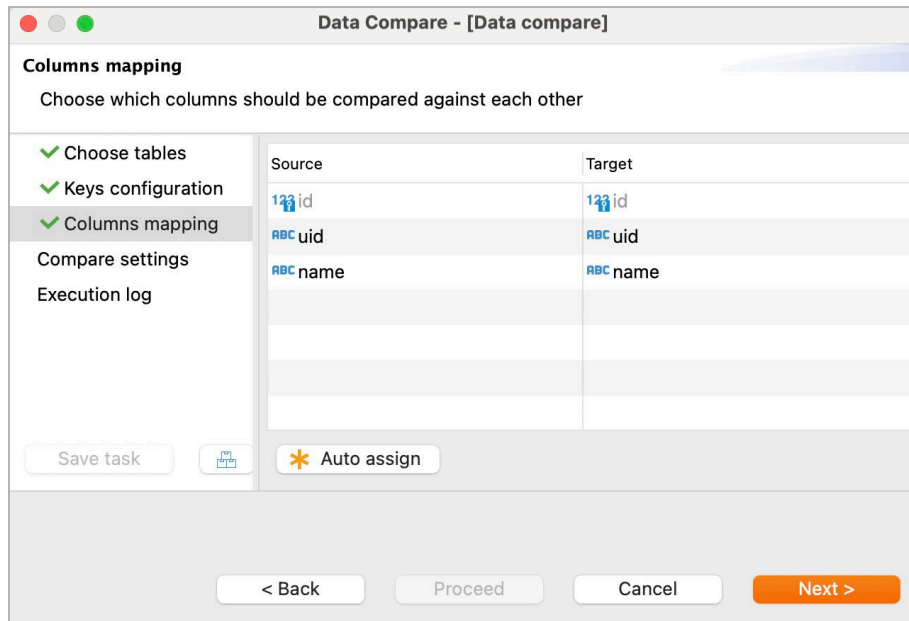


- ID must be unique
- Ideally, it is a primary key
- You need an index for this column if it's not a PK.

# HOW IT WORKS: ADDED AND DELETED



# HOW TO COMPARE: MODIFIED



- Simple check for the modifications only in rows with the same ID
- Check according to the mapping
- Separate selects only for the modified rows

# HOW TO COMPARE: FINAL SCRIPT

Export Changes

Preview of the generated SQL script:

```
DELETE FROM test.table_diff_small_1
WHERE id=0;
UPDATE test.table_diff_small_1
SET uid='valueX',"name"='changed value X'
WHERE id=35;
UPDATE test.table_diff_small_1
SET "name"='changed value Z',uid='valueZ'
WHERE id=135;
INSERT INTO test.table_diff_small_1 (id,"name",uid)
VALUES (1001,'missing row in left table','missing row');
```

Options

Apply changes to:  Source table  Target table



# HOW TO COMPARE: IN APP

Source - test.table_diff_small_1 (PostgreSQL - gis)				Target - test.table_diff_small_2 (PostgreSQL - gis)			
	id	uid	name		id	uid	name
1	0	new row	new row in left tab	1	35	valueX	changed value X
2	35	value35	load tests 35	2	135	valueZ	changed value Z
3	135	value135	load tests 135	3	1,001	missing row	missing row in left

- ✓ No manual work
- ✓ The result is a visual diff and a SQL script
- ✓ Connect directly to any database
- ✓ Database is safe
- ✓ It's fast
- ✗ You need a third-party app
- ✗ Difficulties with the composite keys
- ✗ Requires space on the machine

# WHAT TO CHOOSE?

Conditions	Comparison Method
Tiny table	Compare by eye
Small table, no SQL knowledge	Excel or comparison tools
All data in Postgres, database expert	SQL queries
Data available through FDW, DB expert	SQL queries
Big tables in Postgres	Comparison tools
Data in different databases	Comparison tools

## USEFUL LINKS

EMAIL:	<a href="mailto:tati@dbeaver.com">tati@dbeaver.com</a>
DBEAVER GITHUB:	<a href="https://github.com/dbeaver">github.com/dbeaver</a>
DBEAVER COMMUNITY:	<a href="https://dbeaver.io">dbeaver.io</a>
DBEAVER PRO:	<a href="https://dbeaver.com">dbeaver.com</a>

# THANK YOU