

## Source datasets

 $S_2$  Lineitem $S_1$ : Order

order_id	customer_sk	date
1	11	2024-01-01
2	12	2023-01-02
3	11	2024-01-03
...	...	...

order_id	product_id	quantity	price
1	21	2	20
1	22	1	25
2	23	3	13
3	21	1	10
...	...	...	...

Factorized  
Learning

 $S_3$  Order returns

order_id	product_id	return_quantity
3	21	1
...	...	...

Join

 $T$ : Materialized data

order_id	customer_sk	date	product_id	quantity	price	return_quantity
1	11	2024-01-01	21	2	20	0
1	11	2024-01-01	22	1	25	0
2	12	2024-01-02	23	3	13	0
3	11	2024-01-03	21	1	10	1
...	...	...	...	...	...	...

Learning over  
Materialized data