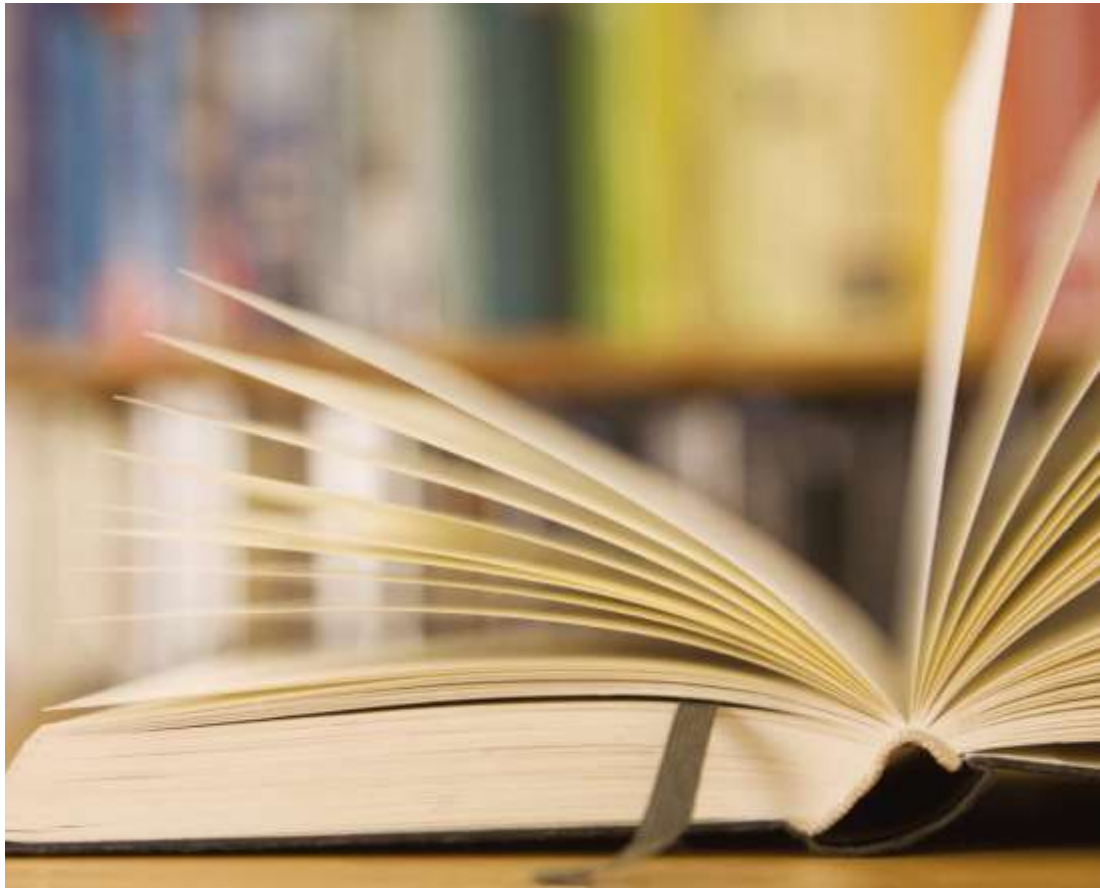
CSCI317 – Database Performance Tuning

Suboptimal clustering



Suboptimal clustering

- The database contains information about the product categories, customer, employees, suppliers, shippers, products, orders, and order details.
- After loading data into the database, the relational tables have the following sizes:

CATEGORY	1 data block
CUSTOMER	200 data blocks
EMPLOYEE	20 data blocks
SUPPLIER	10 data blocks
SHIPPER	10 data blocks
PRODUCT	50 data blocks
ORDERS	300 data blocks
ORDER_DETAIL	900 data blocks

Suboptimal clustering

- We would like to use clustering to improve performance of the following types of queries:
  1) Find full information about the products and categories of the products.
  2) Find order dates of all orders submitted by the customers that belong to a given company.
  3) Find the names of employees who handles the orders submitted in a given year.
  4) Find the names of products supplied by suppliers located phone number.
  5) Find the dates of shipments performed by a shipper that has a given phone number.

Suboptimal clustering

For each one of the queries listed above, find all joins of the relational tables that must be done to process a query.

1) Find full information about the product and categories of the products.

PRODUCT join with CATEGORY

Suboptimal clustering

For each one of the queries listed above, find all joins of the relational tables that must be done to process a query.

2) Find order dates of all orders submitted by the customers that belong to a given company.

ORDERS join with CUSTOMER

Suboptimal clustering

For each one of the queries listed above, find all joins of the relational tables that must be done to process a query.

3) Find the names of employees who handles the orders submitted in a given year.

EMPLOYEE join with ORDERS

Suboptimal clustering

For each one of the queries listed above, find all joins of the relational tables that must be done to process a query.

4) Find the names of products supplied by suppliers located in a given city.

PRODUCT join with SUPPLIER

Suboptimal clustering

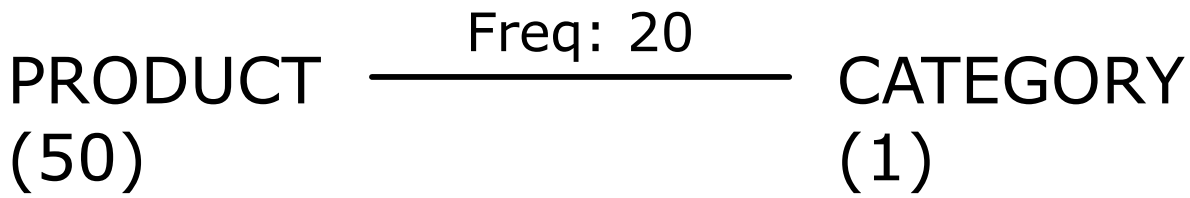
For each one of the queries listed above, find all joins of the relational tables that must be done to process a query.

5) Find the dates of shipments performed by a shipper that has a given phone number.

SHIPPER join with ORDERS

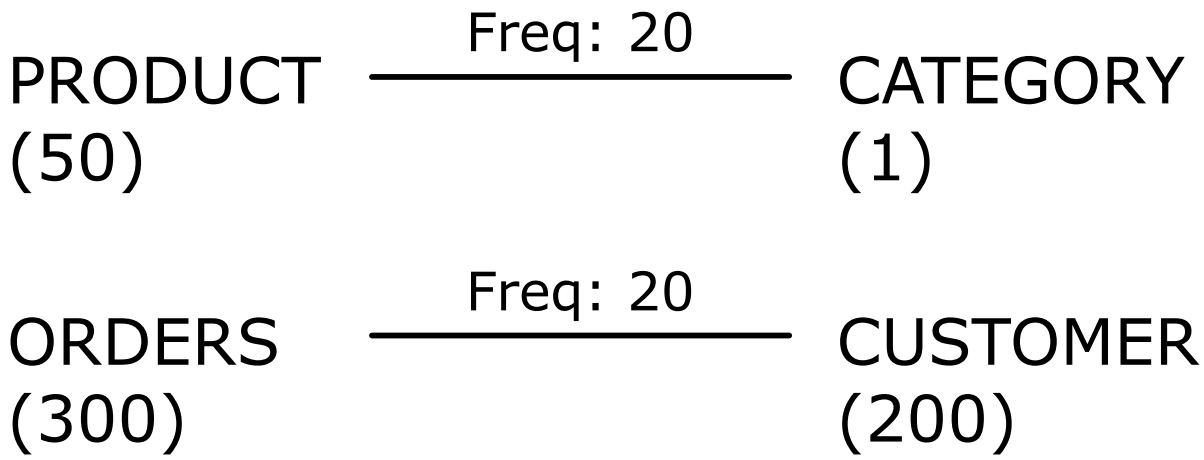
The possible clusters:

PRODUCT join with CATEGORY



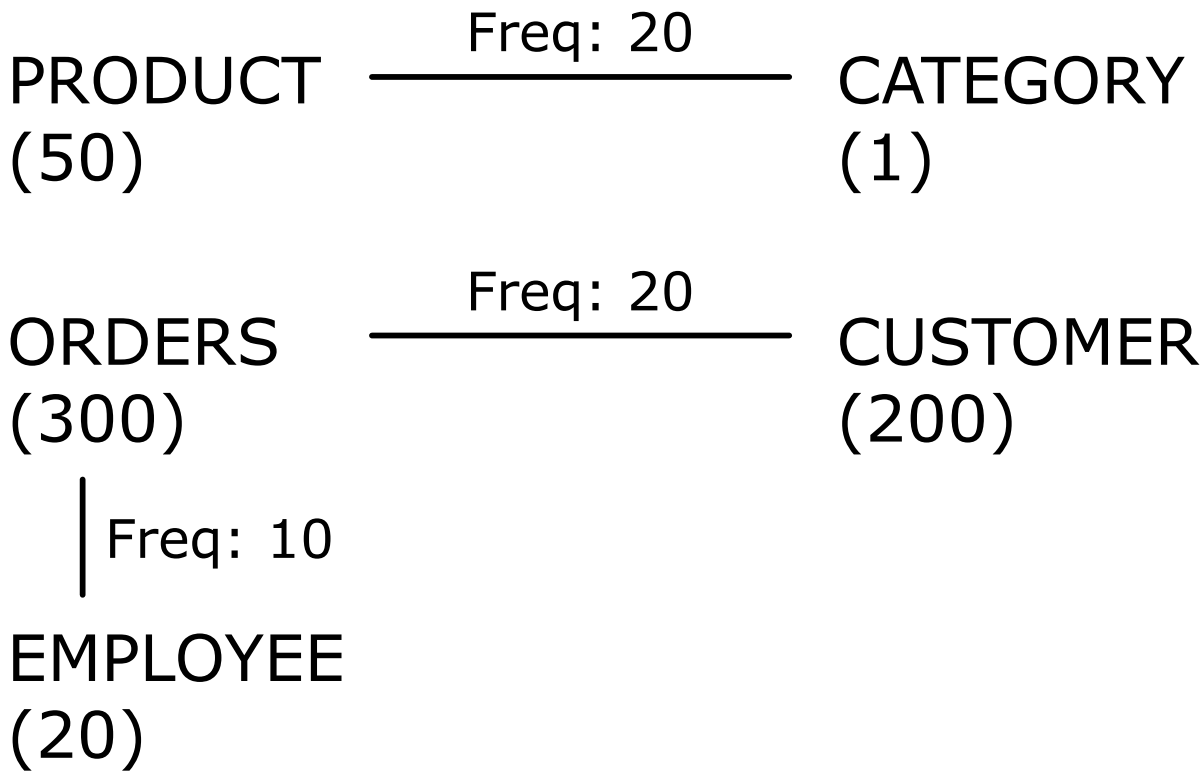
The possible clusters:

ORDERS join with CUSTOMER



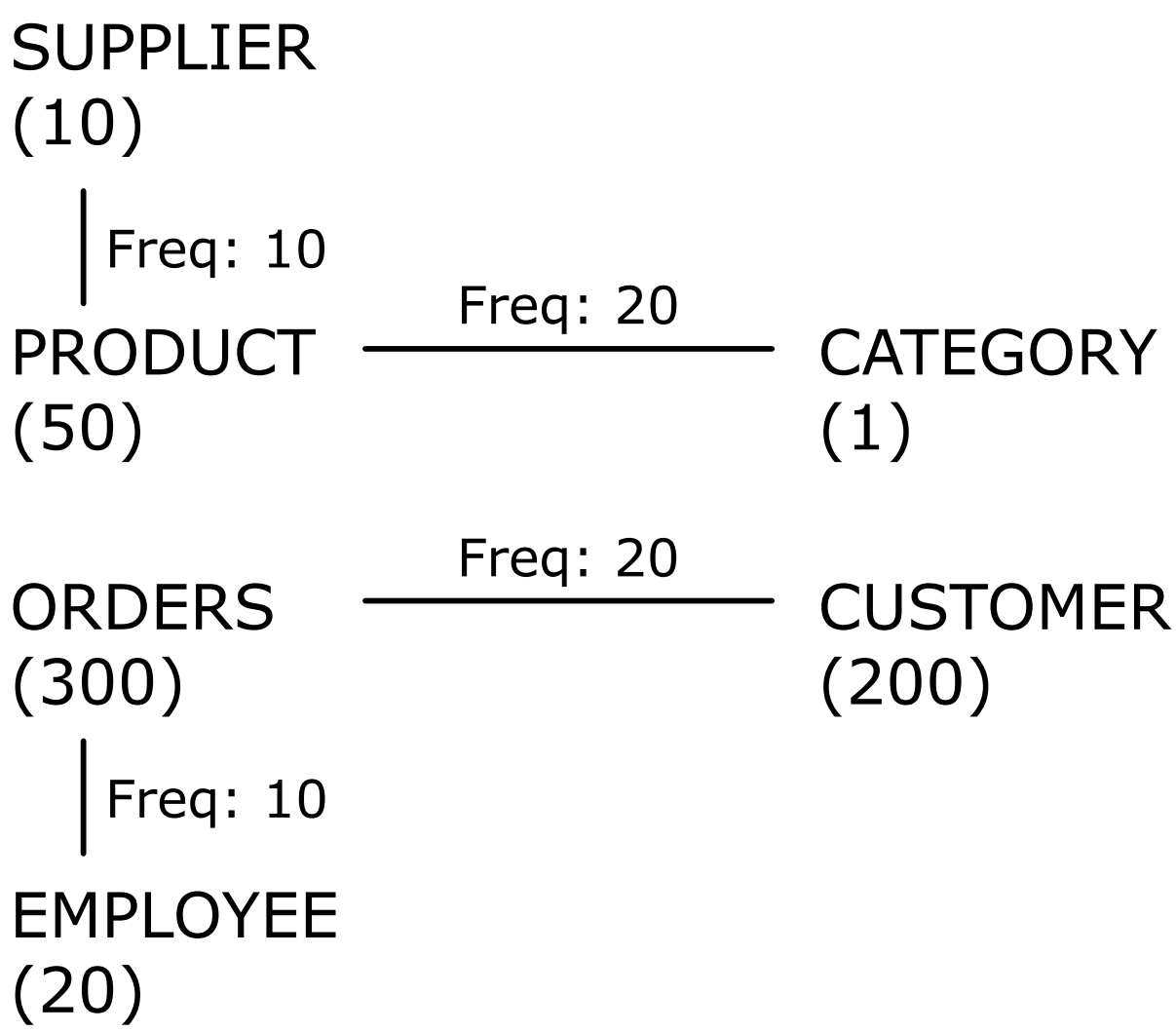
The possible clusters:

EMPLOYEE join with ORDERS



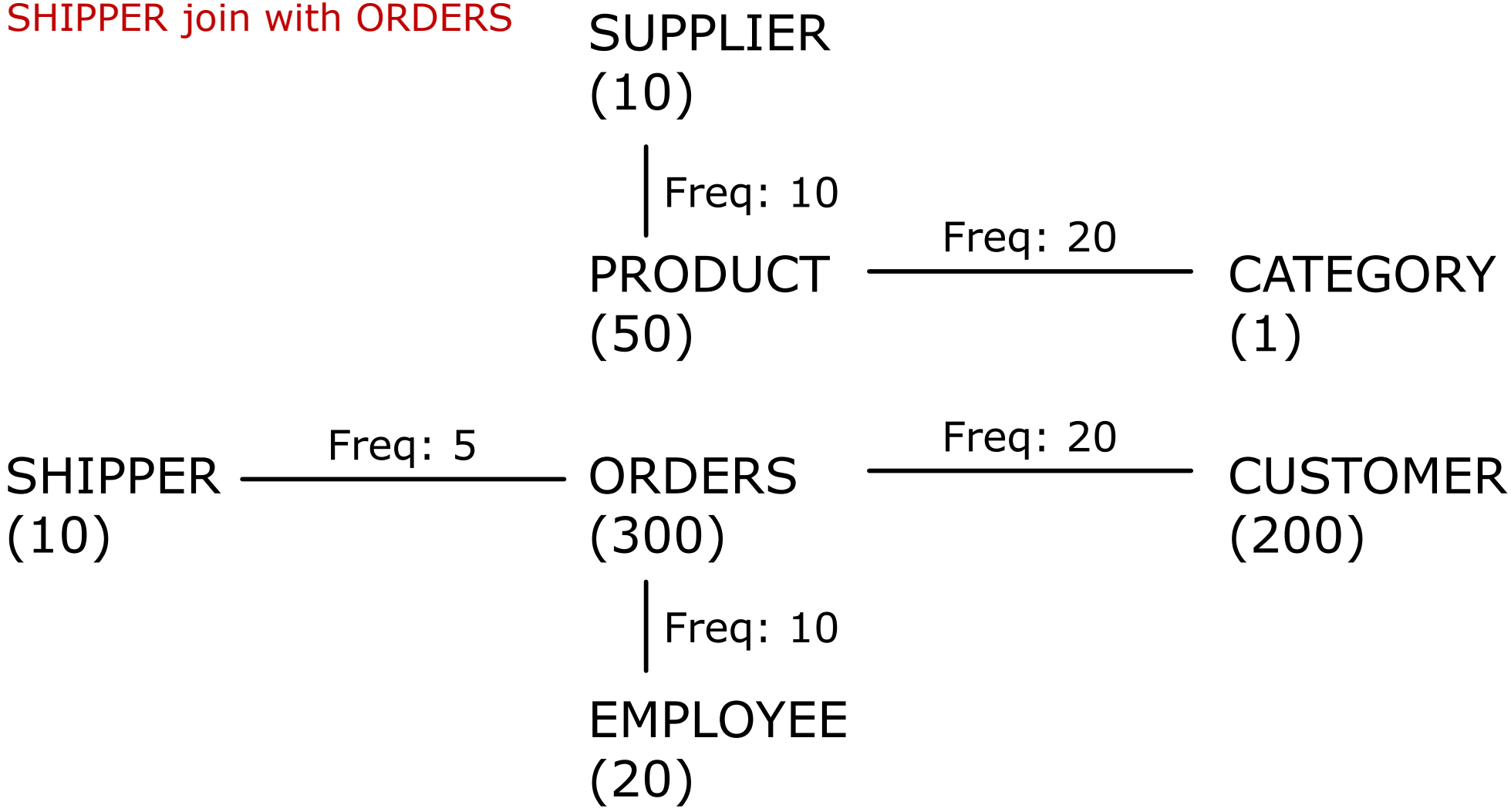
The possible clusters:

PRODUCT join with SUPPLIER

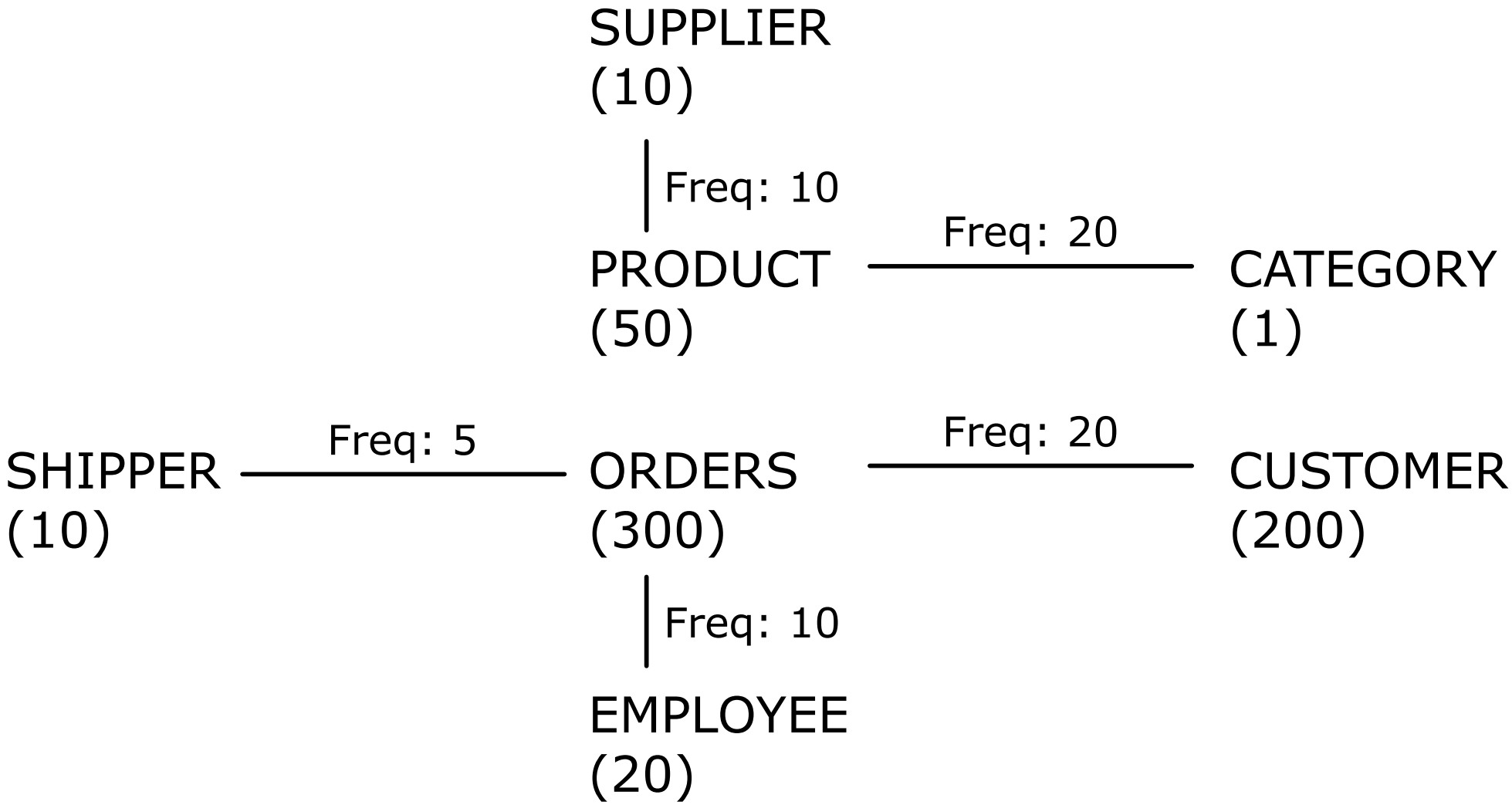


The possible clusters:

SHIPPER join with ORDERS



The possible clusters:



Possible saving:

Cost of join operation: $3 \times (Block_1 + Block_2)$

Cost of sequential scan: $(Block_1 + Block_2)$

$$Possible\ saving\ = \\ [(Cost\ of\ join\ operation) - (Cost\ of\ sequential\ scan)] \\ \times Frequency\ of\ operation$$

Possible saving:

PRODUCT (50) —— Freq: 20 —— CATEGORY (1)

$$Possible\ Saving = [3 \times (Block_1 + Block_2) - (Block_1 + Block_2)] \times Freq$$

$$Possible\ Saving = [3 \times (50 + 1) - (50 + 1)] \times 20$$
$$= 2040 \text{ read block operations}$$

Possible saving:

ORDERS (300) — Freq: 20 — CUSTOMER (200)

$$Possible\ Saving = [3 \times (Block_1 + Block_2) - (Block_1 + Block_2)] \times Freq$$

$$Possible\ Saving = [3 \times (300 + 200) - (300 + 200)] \times 20$$
$$= 20{,}000 \text{ read block operations}$$

Possible saving:

ORDERS (300) —Freq: 10— EMPLOYEE (20)

$$Possible\ Saving = [3 \times (Block_1 + Block_2) - (Block_1 + Block_2)] \times Freq$$

$$Possible\ Saving = [3 \times (300 + 20) - (300 + 20)] \times 10$$
$$= 6{,}400 \text{ read block operations}$$

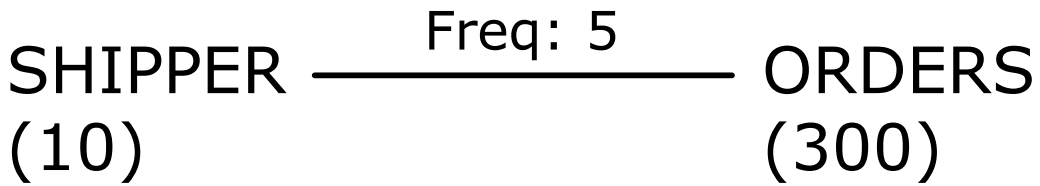
Possible saving:

SUPPLIER (10) — Freq: 10 — PRODUCT (50)

$$Possible\ Saving = [3 \times (Block_1 + Block_2) - (Block_1 + Block_2)] \times Freq$$

$$Possible\ Saving = [3 \times (10 + 50) - (10 + 50)] \times 10$$
$$= 1{,}200 \text{ read block operations}$$

Possible saving:



$$Possible\ Saving = [3 \times (Block_1 + Block_2) - (Block_1 + Block_2)] \times Freq$$

$$Possible\ Saving = [3 \times (10 + 300) - (10 + 300)] \times 5$$
$$= 3{,}100 \text{ read block operations}$$

Possible saving:

$$PossibleSaving_{(Q1)} = [3 \times (50 + 1) - (50 + 1)] \times 20$$
$$= 2040 \text{ read block operations}$$

$$PossibleSaving_{(Q2)} = [3 \times (300 + 200) - (300 + 200)] \times 20$$
$$= 20{,}000 \text{ read block operations}$$

$$PossibleSaving_{(Q3)} = [3 \times (300 + 20) - (300 + 20)] \times 10$$
$$= 6{,}400 \text{ read block operations}$$

$$PossibleSaving_{(Q4)} = [3 \times (10 + 50) - (10 + 50)] \times 10$$
$$= 1{,}200 \text{ read block operations}$$

$$PossibleSaving_{(Q5)} = [3 \times (10 + 300) - (10 + 300)] \times 5$$
$$= 3{,}100 \text{ read block operations}$$

Possible saving:

$$PossibleSaving_{(Q1)} = [3 \times (50 + 1) - (50 + 1)] \times 20$$
$$= 2040 \text{ read block operations}$$

$$PossibleSaving_{(Q2)} = [3 \times (300 + 200) - (300 + 200)] \times 20$$
$$= 20{,}000 \text{ read block operations}$$

$$PossibleSaving_{(Q3)} = [3 \times (300 + 20) - (300 + 20)] \times 10$$
$$= 6{,}400 \text{ read block operations}$$

$$PossibleSaving_{(Q4)} = [3 \times (10 + 50) - (10 + 50)] \times 10$$
$$= 1{,}200 \text{ read block operations}$$

$$PossibleSaving_{(Q5)} = [3 \times (10 + 300) - (10 + 300)] \times 5$$
$$= 3{,}100 \text{ read block operations}$$

Possible saving:

$$PossibleSaving_{(Q1)} = [3 \times (50 + 1) - (50 + 1)] \times 20$$
$$= 2040 \text{ read block operations}$$

$$PossibleSaving_{(Q2)} = [3 \times (300 + 200) - (300 + 200)] \times 20$$
$$= 20{,}000 \text{ read block operations}$$

$$PossibleSaving_{(Q3)} = [3 \times (300 + 20) - (300 + 20)] \times 10$$
$$= 6{,}400 \text{ read block operations}$$

$$PossibleSaving_{(Q4)} = [3 \times (10 + 50) - (10 + 50)] \times 10$$
$$= 1{,}200 \text{ read block operations}$$

$$PossibleSaving_{(Q5)} = [3 \times (10 + 300) - (10 + 300)] \times 5$$
$$= 3{,}100 \text{ read block operations}$$

The possible clusters: