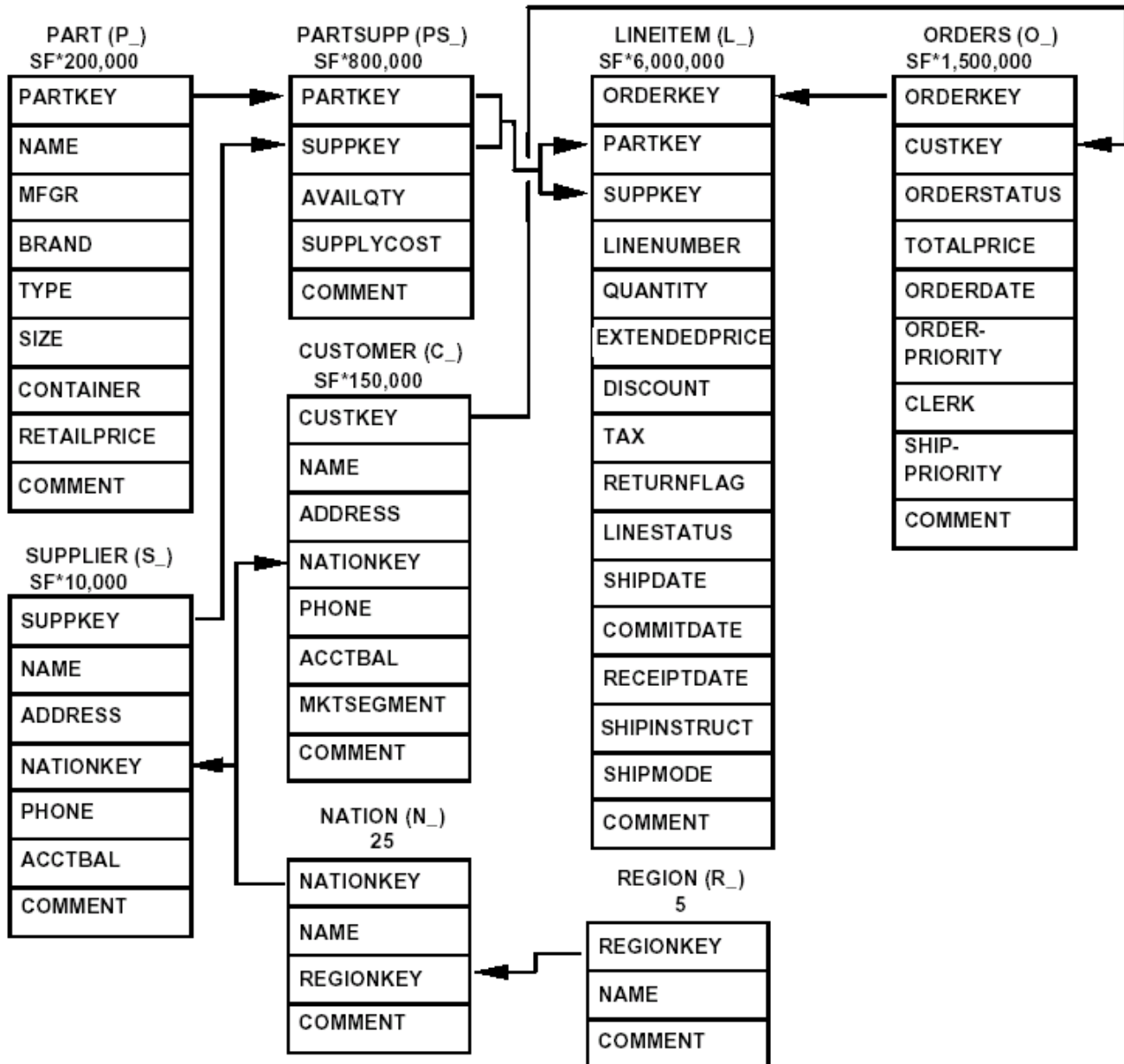


Relational Algebra: Aggregation

Database Schema



Legend:

- The parentheses following each table name contain the prefix of the column names for that table;
- The arrows point in the direction of the one-to-many relationships between tables;
- The number/formula below each table name represents the cardinality (number of rows) of the table. Some are factored by SF, the Scale Factor, to obtain the chosen database size. The cardinality for the LINEITEM table is approximate (see Clause 4.2.5).

Problems

1. Find the 10 most expensive orders that contained at least 200 items.
Note: Take into consideration each LineItem's quantity field.
2. Find the 5 largest volumes (\$-wise) shipped between any two countries.
3. Find the percentage of customers from each nation.

Answers

1.

```
τ O.totalPrice DESC
|
P qty = SUM(L.quantity)
|
π O.orderKey, O.totalPrice, SUM(L.quantity)
|
σ sum(L.quantity) > 200
|
Υ O.orderKey, sum(L.quantity)
|
⋈
O.orderKey = L.orderKey
/      \
Orders O      LineItem L
```

```
SELECT O.orderKey, O.totalPrice, SUM(L.quantity) as qty
FROM Orders O, LineItem L
WHERE O.orderKey = L.orderKey
GROUP BY O.orderKey
HAVING SUM(L.quantity) > 200
ORDER BY O.totalPrice DESC
LIMIT 10;
```

orderKey	totalPrice	qty
52965	466001.28	271
29158	439687.23	305
44707	431771.98	279
59106	430619.75	276
6882	422359.65	303
57376	411255.46	240
39456	409770.83	266
17571	408345.74	249
39620	406938.36	272
35460	405742.27	242

2.

```

      τ volume DESC
      |
P volume = sum(L.extendedPrice * (1-L.discount))
      source = N1.name
      destination = N2.name
      |
      π
      sum(L.extendedPrice * (1-L.discount)),
      N1.name, N2.name
      |
Υ S.nationKey, C.nationKey , sum(L.extendedPrice * (1-L.discount))
      |
      σ N1.nationKey != N2.nationkey
      |
      ⋈
      L.orderKey = O.orderKey
      /          \
      ⋈          ⋈
S.suppKey = L.suppKey      O.custKey = C.custKey
      /          \          /          \
      ⋈          LineItem L      Orders O      ⋈
S.nationKey = N1.nationKey      C.nationKey = N2.nationKey
      /          \          /          \
Supplier S      Nation N1      Customer C      Nation N2
```

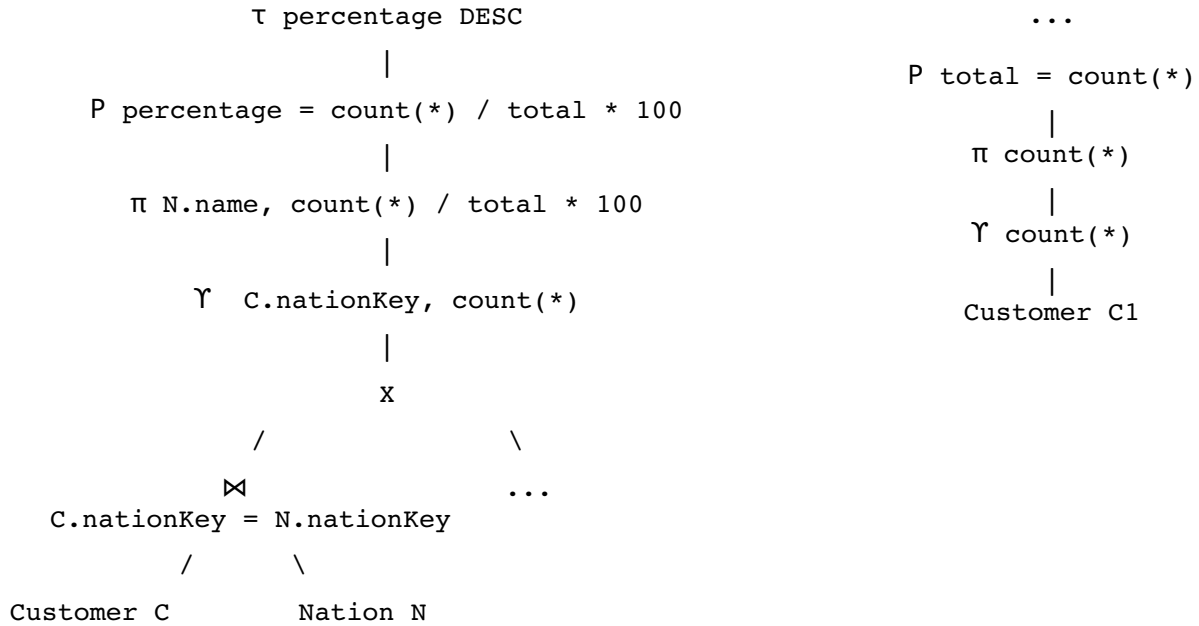
```

SELECT
    sum(L.extendedPrice * (1-L.discount)) as volume,
    N1.name as source, N2.name as destination
FROM Customer C, Supplier S, LineItem L, Orders O, Nation N1, Nation N2
WHERE
    C.custKey = O.custKey AND
    S.suppKey = L.suppKey AND
    O.orderkey = L.orderKey AND
    S.nationKey = N1.nationKey AND
    C.nationKey = N2.nationKey AND
    N1.nationKey != N2.nationKey
GROUP BY S.nationKey, C.nationKey
ORDER BY volume DESC
LIMIT 5;

```

volume	source	destination
8277721.06	UNITED STATES	CANADA
8191370.72	UNITED STATES	ROMANIA
8051955.22	CHINA	CANADA
7918617.50	UNITED STATES	EGYPT
7883487.67	UNITED STATES	IRAN

3.



```

SELECT N.name, count(*) / (SELECT count(*) FROM Customer) * 100 as percentage
FROM Customer C, Nation N
WHERE C.nationKey = N.nationKey
GROUP BY C.nationKey
ORDER BY percentage DESC;
  
```

name	percentage
IRAN	4.8000
MOROCCO	4.8000
CANADA	4.6000
BRAZIL	4.5333
SAUDI ARABIA	4.4667
JAPAN	4.4667
INDONESIA	4.4000
EGYPT	4.4000
ROMANIA	4.2667
MOZAMBIQUE	4.1333
ALGERIA	4.0667
INDIA	4.0000
RUSSIA	3.9333
ARGENTINA	3.9333
VIETNAM	3.8667
CHINA	3.8667
IRAQ	3.8667
ETHIOPIA	3.8000
GERMANY	3.8000
PERU	3.7333
...	...