Task 5 (3 marks)
An objective of this task is to find the best distribution of relational tables over the persistent storage devices.

Assume, that to avoid the conflicts with the accesses to the relational tables of TPC-HR sample database we would like to distribute the relational tables over two different persistent storage devices. Then the relational tables that are joined together can be simultaneously read from two or more persistent storage devices. Do not worry if your system does not have persistent storage devices. We shall simulate the drives through two different tablespaces `DRIVE_C` and `DRIVE_D`. You do not have to create the tablespaces. To find out, which relational tables should be located on each device we shall consider the following queries.

(i) Find the extended prices of items (column `L_EXTENDEDPRICE`) that are available in at least a given quantity (column `PS_AVAILQTY`).
(ii) Find the total prices (column `O_TOTALPRICE`) of the orders that have been submitted by a customer that belongs to nation with a given name (column `N_NAME`).
(iii) Find the names of parts (column `P_NAME`) available in at least a given quantity (column `PS_AVAILQTY`).
(iv) Find the names of suppliers (column `S_NAME`) who live in a region with given region name (column `R_NAME`).

Note, that the prefixes of the column names indicate the relational tables the columns are located at. For example, `R_NAME` denotes a column in a relational table `REGION`.

Analyze the queries listed above and find which relational tables are used by each query and distribute the relational tables over the hard drives simulated by the tablespaces `DRIVE_C` and `DRIVE_D` such, that the relational tables used by the same query are located on the different hard drives. Such approach reduces the total number of conflicts when accessing the persistent storage devices and it speeds up the query processing. If it is impossible to distribute the relational tables used by the same application on the different hard drives then try to minimize the total number of conflicts. You do not need to worry about distribution of indexes used for processing of the queries.

Create a document `solution5.pdf` that contains the following information.

(1) For each one of the queries listed above find what relational table are used by a query and draw an undirected hypergraph such that each one of its hyperedges contains the names of tables used by one query. The names of tables are the nodes of the hypergraph.

(2) Use the hypergraph created in the previous step to find distribution of the relational tables over the persistent storage devices `DRIVE_C` and `DRIVE_D` such, that the relational tables used by the same query are located on the different persistent storage devices. If it is impossible to do it locate smaller relational tables on the same device

and larger relational tables on different devices. Include information which relational table and which index is assigned to which device in a document `solution5.pdf`.

Hint
You can find a definition and visualization of an undirected hypergraph at:
`https://en.wikipedia.org/wiki/Hypergraph`

Deliverables
A file `solution5.pdf` that contains a hypergraph created in step (1) and information about relational tables assigned to the persistent storage devices. You are allowed to use any line drawing tool to draw a hypergraph. A scanned/photographed copy of a neat hand drawing is also acceptable.

Solution

(i) Find the extended prices of items (column `L_EXTENDEDPRICE`) that are available in at least a given quantity (column `PS_AVAILQTY`).

```
LINEITEM JOIN PARTSUPP
```

(ii) Find the total prices (column `O_TOTALPRICE`) of the orders that have been submitted by a customer that belongs to nation with a given name (column `N_NAME`).

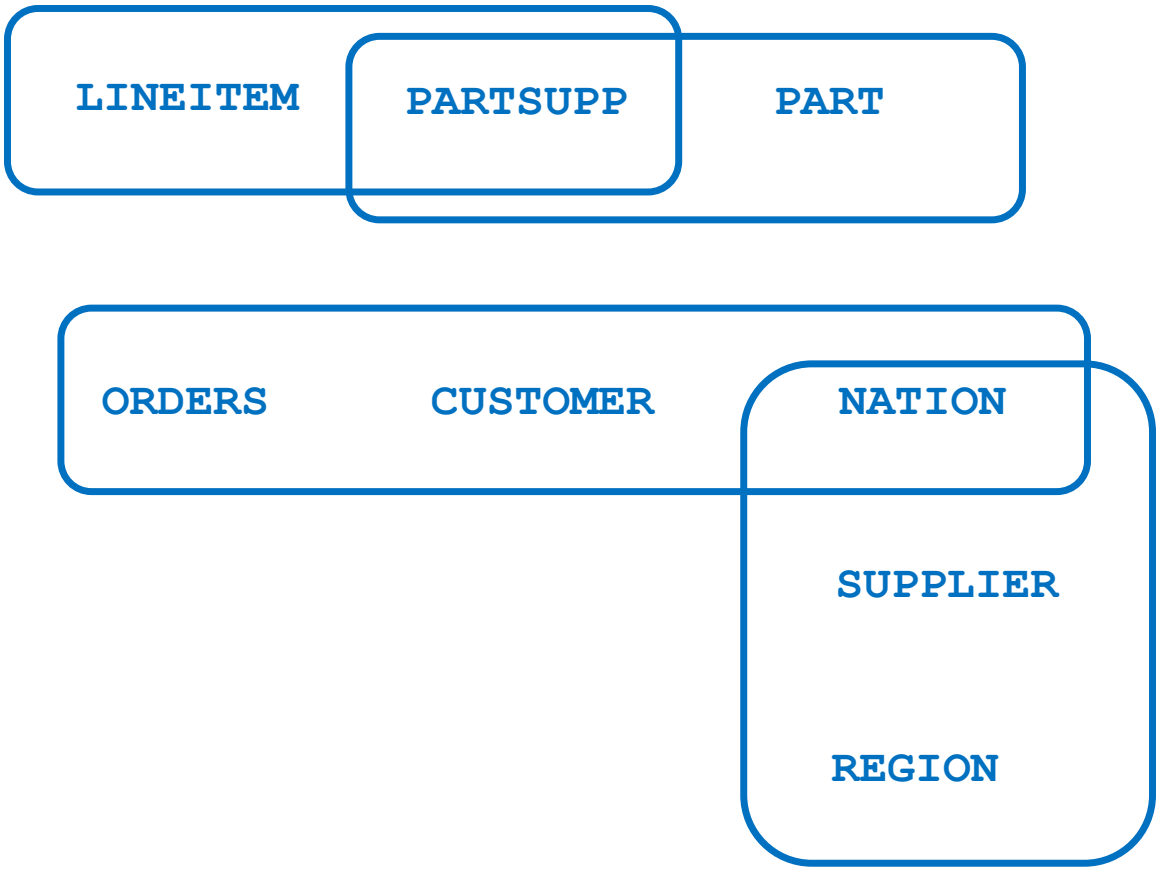
```
ORDERS JOIN CUSTOMER JOIN NATION
```

(iii) Find the names of parts (column `P_NAME`) available in at least a given quantity (column `PS_AVAILQTY`).

```
PART JOIN PARTSUPP
```

(iv) Find the names of suppliers (column `S_NAME`) who live in a region with given region name (column `R_NAME`).

```
SUPPLIER JOIN NATION JOIN REGION
```




```
DRIVE_C: LINEITEM, PART, CUSTOMER, NATION

DRIVE_D: PARTSUPP, ORDERS, SUPPLIER, REGION
```

End of sample solution