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 when accessing the relational tables of TPC-HR sample database we would like to distribute the relational tables over three 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 the persistent storage devices. We shall simulate the persistent storage devices as three different tablespaces `DRIVE_C`, `DRIVE_D` and `DRIVE_E`. You do not have to create the tablespaces. The tablespaces are already created such that each one is located on a different persistent storage device.

Consider the following queries.

(i) *Find the names and addresses (`C_NAME`, `C_ADDRESS`) of customers located in a region with a given name (`R_NAME`).*
(ii) *Find the names of parts (`P_NAME`) included in the orders that have a given shipment date (attribute `L_SHIPDATE`) and a given supply cost (`PS_SUPPLYCOST`).*
(iii) *Find the region names (`R_NAME`) of suppliers that have an account balance (`S_ACCTBAL`) lower than a given value.*
(iv) *Find the names of customers (`C_NAME`) who ordered at least one part shipped in a given year (`L_SHIPDATE`).*
(v) *Find the total number of customers per each nation (`N_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`.

An objective of this task is to analyze the queries listed above to find the relational tables used by each query. Then, to distribute the relational tables over the persistent storage devices simulated by the tablespaces `DRIVE_C`, `DRIVE_D` and `DRIVE_E` such, that each relational table used by the same query is located on a different persistent storage device.

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 persistent storage devices then try to minimize the total number of conflicts. You do not need to worry about the 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 tables 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.

(i) *Find the names and addresses (`C_NAME`, `C_ADDRESS`) of customers located in a region with a given name (`R_NAME`).*

```
CUSTOMER join NATION join REGION
```

(ii) *Find the names of parts (`P_NAME`) included in the orders that have a given shipment date (attribute `L_SHIPDATE`) and a given supply cost (`PS_SUPPLYCOST`).*

```
PART join PARTSUPP join LINEITEM
```

(iii) *Find the region names (`R_NAME`) of suppliers that have an account balance (`S_ACCTBAL`) lower than a given value.*

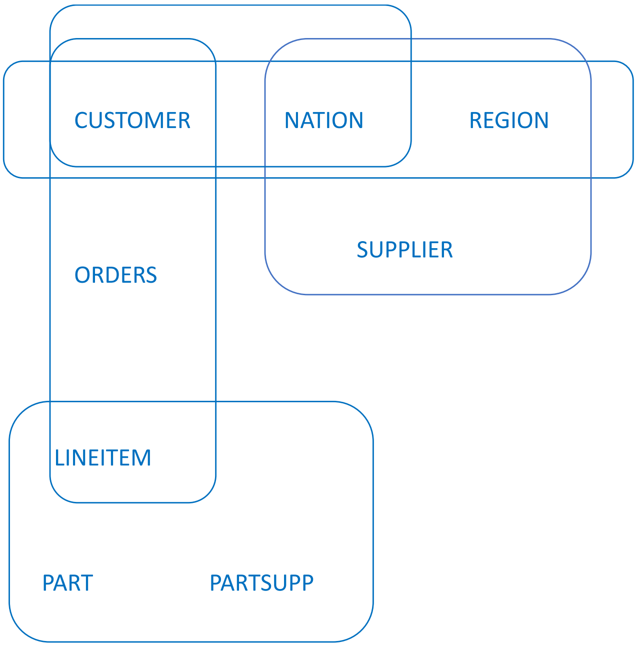
```
REGION join NATION join SUPPLIER
```

(iv) *Find the names of customers (`C_NAME`) who ordered at least one part shipped in a given year (`L_SHIPDATE`).*

```
CUSTOMER join ORDERS join LINEITEM
```

(v) *Find the total number of customers per each nation (`N_NAME`).*

```
CUSTOMER join NATION
```



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

```
DRIVE_C: CUSTOMER, SUPPLIER, PART
DRIVE_D: NATION, ORDERS, PARTSUPP
DRIVE_E: REGION, LINEITEM
```

Include a drawing of a hypergraph obtained from a step (1) and information which relational table is assigned to which device obtained from a step (2) 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 drawing of a hypergraph obtained from a step (1) and information which relational table is assigned to which device obtained from a step (2). 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.
