Assignment2 TPC-H Schema

There are several search problems on some related table named *TPC-H Schema*. You have to choose some of the problems and give out solution since the sum of the points of all questions is greater than 100. In general, harder problem means more points.

The following picture shows the structure of these tables.

Note that the first attribute is primary key of that table (example: PARTKEY is primary key of PART)

PART (P_) PARTSUPP (PS_) LINEITEM (L_) ORDERS (O_) SF*800,000 SF*1,500,000 SF*200,000 SF*6,000,000 PARTKEY PARTKEY ORDERKEY ORDERKEY NAME SUPPKEY PARTKEY CUSTKEY MFGR SUPPKEY ORDERSTATUS AVAILQTY BRAND SUPPLYCOST LINENUMBER TOTALPRICE TYPE QUANTITY ORDERDATE COMMENT ORDER-SIZE EXTENDEDPRICE CUSTOMER (C_) PRIORITY SF*150,000 DISCOUNT CONTAINER CLERK CUSTKEY RETAILPRICE TAX SHIP-NAME PRIORITY RETURNFLAG COMMENT ADDRESS COMMENT LINESTATUS SUPPLIER (S_) NATIONKEY SHIPDATE SF*10,000 PHONE SUPPKEY COMMITDATE ACCTBAL NAME RECEIPTDATE MKTSEGMENT ADDRESS SHIPINSTRUCT COMMENT NATIONKEY SHIPMODE PHONE NATION (N_) COMMENT ACCTBAL NATIONKEY REGION (R_) COMMENT NAME REGIONKEY REGIONKEY NAME COMMENT COMMENT

Figure 2: The TPC-H Schema

Task1

Create a user with your name and a database named tpc_h, import all tables in the file folder *data*, and list all tables in your database. (5 points)

Task2

- a) Display name,address,nationkey of the supplier whose suppkey is 717. (5 points)
- b) For each **ship priority**, find the total number of **orders** that are ordered in 1992-1993. Display the ship priority and the corresponding total number of orders in descending order. (8 points)
- c) Find the **suppliers** that can supply the **part** whose key is 360. Display their names, phones, nations and regions.(8 points)
- d) Display the **keys** and **names** of **suppliers** who supplied more **parts** in 1993 than he did in 1992, and the **number of parts** he supplied in 1992 and 1993. Notice that you need to show the number of part he supplied in 1992 and 1993 separately in you answer.(15 points)
- e) Display the keys of orders who has suppliers come from Europe. (8 points)

f)	Display the customer who is not European .(8 points)
g)	Display average quantity of lineitems which are ordered in November 1997 for each ship mode except "AIR" and "MAIL". Show the results in
	ascending order of the average quantity.(15 points)
h)	Show the total number of customers in each region . (8 points)
i)	Display all the info (partkey,name,,coment)of the top 5 parts with the largest
	size.(5 points)
j)	Find the customer whose phone is 25-XXX-XXXX (5 points)
k)	Find the lineitem which has the maximum extended price and is ordered before 1997. Display the value of this maximum extended price.(10 points)

Task3

Use b_tree or hash index to accelerate problem a), j) and k) in Task2

Show your acceleration and try to compare the efficiency of different indexes.(20 points)