

M2 – IAD - BDMD

Requêtes décisionnelles avec TPC-H

L'objectif de ce TD/TME est de savoir exprimer en SQL des requêtes d'aide à la décision.

Préparation

Voir la spécification du benchmark TPC-H. Comprendre le schéma E/A de l'application donné en annexe.

Exercice 1 : Requêtes TPC-H

Soit le schéma relationnel d'une application de gestion :

Part (p_partkey, p_name, p_mfgr, p_brand, p_type, p_size, p_container, p_retailprice, p_comment)

mfgr est le nom de groupe du fabricant, *brand* est la marque du fabricant.

Supplier (s_suppkey, s_name, s_address, s_nationkey, s_phone, s_acctbal, s_comment)

s_acctbal est le solde du fournisseur

PartSupp (ps_partkey, ps_suppkey, ps_availqty, ps_supplycost, ps_comment)

Customer (c_custkey, c_name, c_address, c_nationkey, c_phone, c_acctbal, c_mktsegment, c_comment)

mktsegment : le type de marché (auto, meuble,...)

Orders (o_orderkey, o_custkey, o_orderstatus, o_totalprice, o_orderdate, o_orderpriority, o_clerk, o_shippriority, o_comment)

o_orderstatus: état de la commande : non traitée (O), en cours (P), terminée (F).

o_orderpriority 5 niveaux de priorité : de urgent à faible.

Linitem (l_orderkey, l_partkey, l_suppkey, l_linenum, l_quantity, l_extendedprice, l_discount, l_tax, l_returnflag, l_linestatus, l_shipdate, l_commitdate, l_receiptdate, l_shipinstruct, l_shipmode, l_comment)

l_linenum : un n° de 1 à 7

l_returnflag est l'état de réception d'un article: non reçu(N), accepté(A), ou retour à l'expéditeur(R),

l_linestatus est l'état d'expédition d'un article : à envoyer(O) ou envoyé(F)

l_shipdate: expédition dans un délai max de 6 mois après la commande,

l_commitdate: paiement par le client dans un délai max de 3 mois après la commande,

l_receiptdate : réception dans un délai max d'un mois après l'expédition.

La **clé** est (orderkey,linenum)

Nation (n_nationkey, n_name, n_regionkey, n_comment)

Region (r_regionkey, r_name, r_comment)

une région est une zone formée de plusieurs pays voisins (\approx continent)

Les attributs soulignés définissent une clé primaire. Deux attributs de même nom sont liés par une contrainte d'intégrité référentielle. L'attribut clé étrangère fait référence à l'attribut clé primaire. Par exemple : Supplier.nationkey fait référence à Nation.nationkey, PartSupp.partkey fait référence à Part.partkey (ceci seulement pour les attributs dont le suffixe est "key").

Les attributs suivants ont une valeur positive : toutes les clés ainsi que les attributs size, retailprice, availqty, supplycost, totalprice, quantity, extendedprice, tax. L'attribut discount est un pourcentage défini dans [0,1]. Les dates sont ordonnées telles que : orderdate \leq shipdate \leq receiptdate et orderdate \leq commitdate.

On donne la cardinalité des relations : card(Part) = 200 000 * sf, card(Orders) = 1 500 000 * sf

Concernant la jointure entre Customer et Order : 1/3 des clients n'a pas de commande. Les commandes sont associées avec 2/3 des clients, choisis aléatoirement.

Question 1

Ecrire en SQL les requêtes suivantes.

R1: Résumé des prix

Quels sont les articles commandés, livrés avant la date d (inclusive), regroupés selon leur état d'expédition et de réception ? Le résultat de la requête contient le nombre d'articles, le prix total HT (avant et après la réduction) et TTC, la quantité moyenne, le prix moyen, la réduction moyenne et le nombre de commandes.

R2:Fournisseur proposant les meilleurs prix

Soient une région R , et les produits de taille S et de type T . Quels sont les fournisseurs de la région R proposant les meilleurs prix pour les produits de type T et de taille S ? Trier le résultat selon le solde du fournisseur (ordre décroissant). Donner le solde et le nom du fournisseur, le nom de son pays, le numéro et la marque du produit, l'adresse et le n° de téléphone du fournisseur.

R3: Commandes à expédier en priorité

Pour un type de marché S , quelles sont les commandes, non encore expédiées à la date d , les plus importantes en terme de montant restant à encaisser (i.e., la somme du prix HT, après réduction, des articles non expédiés de la commande) ? Trier le résultat par ordre décroissant de prix.

R4: Efficacité du système de priorité des commandes.

On suppose qu'un client est davantage satisfait s'il reçoit sa commande avant de la payer. Autrement dit, une commande ne satisfait pas le client si elle est reçue après le paiement.

Pour diminuer le nombre de clients insatisfaits, l'entreprise modifie la chaîne de traitement des commandes pour mettre en place un système de priorité. Afin de vérifier l'efficacité du système de priorité, on veut déterminer le nombre de commandes non satisfaisantes en fonction de la priorité des commandes.

Pour un trimestre donné (de la date d à $d + 3$ mois), combien de commandes ont au moins un article reçu après la date de paiement ? Donner la priorité et le nombre de commandes.

R5: Volume du marché intérieur

Pour une année donnée (de d à $d + 1$ an) et pour chaque pays de la région R , donner le montant total des commandes intérieures, i.e., pour lesquelles le client et le fournisseur sont dans ce pays. Trier le résultat par montant décroissant.

R6 .Impact des promotions sur le chiffre les revenus

Soient une quantité Q et une réduction R . Pour une année donnée (de d à $d + 1$ an), quel est le montant rapporté par les commandes d'articles en quantité inférieure à Q et qui ont bénéficié d'une réduction de $R \pm 1\%$.

R7: Volume des échanges internationaux

La négociation des contrats avec les sociétés de transport de marchandise, s'appuie sur le montant des transactions entre 2 pays.

Pour une période de n années, quel est le montant annuel des commandes qui ont nécessité une livraison entre les pays $N1$ et $N2$. Donner le pays du fournisseur, le pays de livraison, l'année et le montant. Trier le résultat selon le pays du fournisseur, le pays de livraison et l'année.

R9: Bénéfice ventilé par année, pays et produit

Pour une gamme de produits dont le nom contient le mot P , quel est le bénéfice total pour chaque année et chaque pays. Le bénéfice est la différence entre le prix de vente au client et le prix d'achat au fournisseur.

R11: Articles en stock les plus importants

Le stock d'un pays est l'ensemble des stocks des fournisseurs dans ce pays. La valeur d'un article est

$$\text{valeur} = \text{prix fournisseur} * \text{quantité}$$

Soit une fraction f entre 0 et 1, un article est important sa valeur représente plus de f % de la valeur totale du stock.

Pour un pays P , quels sont les articles importants ? Donner le numéro d'article et sa valeur. Trier le résultat par ordre décroissant de valeur

R13 Distribution des clients selon leur nombre de commandes

Une commande est de type T si son attribut *comment* contient les mots $M1$ puis $M2$. Soit n le nombre de commandes faites par un client (n pouvant être égal à 0), sans tenir compte des commandes de type T .

Pour chaque valeur de n , donner le nombre de clients correspondant.

R21: Fournisseurs retardataires

Cette requête indique les fournisseurs d'un certain pays P qui ne sont pas capables de livrer des articles pour que le client les reçoive avant la date de paiement.

Un fournisseur est retardataire s'il existe au moins un commande déjà envoyée telle que les articles de cette commande ont été livrés par plus d'un fournisseur mais que lui seul n'a pas pu livrer un article de telle sorte qu'il arrive chez le client avant la date de paiement.

Donner le nom des fournisseurs retardataires et le nombre de commandes qu'ils ont retardées.

R22: Opportunité de vente

Pour chaque pays d'un ensemble de pays, combien de clients ont un solde créditeur supérieur à la moyenne et n'ont acheté aucun produit depuis au moins 7 ans ?

Un client appartient à un ensemble de pays si le préfixe de son numéro de téléphone est parmi les valeurs $\{P1, \dots, Pn\}$.

Exercice 2 : Comparaison de TPC-H avec TPC-C

Soit le schéma TPC-C d'une application transactionnelle

Warehouse (wid, name, street, city, state, zip, tax, ytd)

ytd est le solde courant

District (did, wid, name, street, city, state, zip, tax, ytd, next_oid)

10 did par wid

next_oid est le prochain n° de commande disponible

Customer (cid, did, wid, first, middle, last, street, city, state, zip, phone, since, credit, credit_limit, discount, balance, ytd, payment_cnt, deliv_cnt, data)

3000 clients par district

credit vaut 'GC' (good) pour 10% des clients 'BC' (bad) pour les autres

discount entre 0 et 50%

History (cid, c_did, c_wid, did, wid, date, amount, data), sans clé

le client [cid, c_did, c_wid] paie le magasin [did, wid]

NewOrder (oid, did, wid)

Order (oid, did, wid, cid, entry_date, carrier, ol_cnt, all_local)

ol_cnt: nb de lignes d'une commande

all_local vaut 1 si tous les articles de la commande viennent du même endroit, 0 sinon.

Orderline (oid, did, wid, olnum, item, supply_wid, delivery, quantity, amount, info)

olnum : n° de ligne dans [1,15]

Item (item, imageid, name, price, data)

100 000 articles

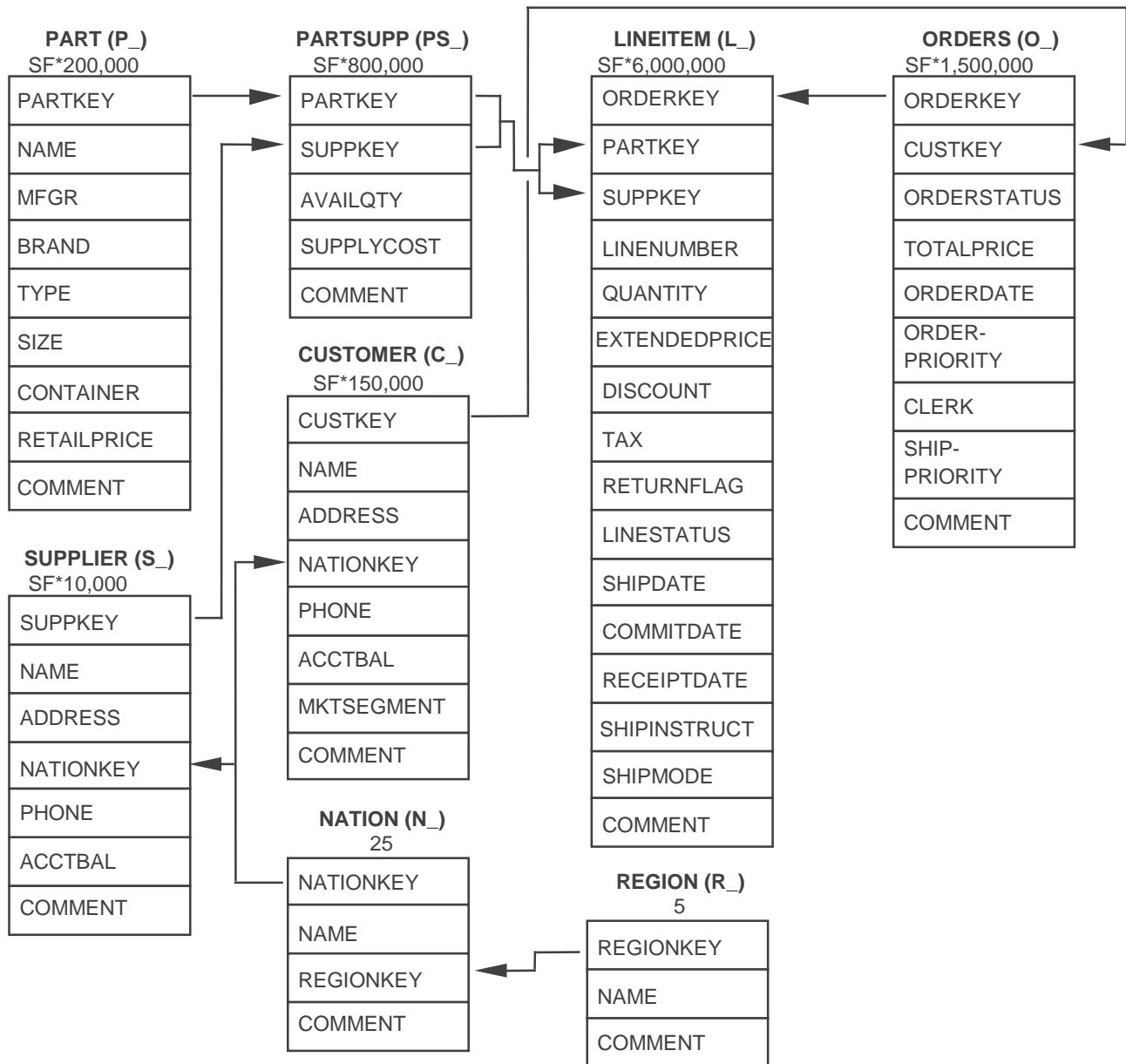
Stock (item, wid, quantity, d1, ..., d10, ytd, order_cnt, remote_cnt, data)

100 000 par wid

Peut-on exprimer les requêtes TPC-H sur le schéma TPC-C ?

Reprendre les requêtes de l'exercice précédent, écrire en SQL, lorsque cela est possible une requête TPC-C ayant une définition proche.

Annexe : le schéma entités-associations de l'application TPC H



Rmq : Les noms des attributs sont tous préfixés (exemple : préfixe **P_** pour les attributs de la relation Part).