

Création d'une base de données de clients

Cahier de charges

Sur base du schéma de base de données présenté ci-dessous et analysé au cours, concevoir et réaliser un programme en C sous CodeBlocks en mode console, géré par des menus simples. La base de données physique sera constituée d'un fichier binaire unique contenant toutes les données et les index nécessaires. Un fichier de trace listera les opérations réalisées sur la base.

Le logiciel devra supporter:

- La création de la base de données
- L'importation des 6 tables fournies sous forme de fichier .csv
- L'exportation de ces 6 tables sous forme de fichier .csv
- La création des index nécessaires aux requêtes précisées
- Des écrans permettant de lister l'information par **recherche séquentielle** pour les tables de codes
- Des écrans permettant d'afficher par **recherche dichotomique**
 - Une personne donnée par sa clé primaire
 - Une compagnie donnée par sa clé primaire
- Au moins un écran permettant d'afficher via une **liste chaînée en mémoire**
 - Les personnes travaillant pour une compagnie donnée par sa clé primaire
 - Les compagnies appartenant à un groupe donné
- Au moins un écran utilisant un **index binaire sur disque** dans la recherche
 - Les personnes travaillant pour une compagnie donnée par sa clé primaire
 - Les personnes dont le nom commence par une chaîne donnée
- Les listes chaînées pourront être triées ascendant ou descendant
- Deux rapports répondants aux requêtes précisées ci-dessous

Pour réduire la portée du projet, celui-ci sera essentiellement «read-only», une fois les données importées dans la base.

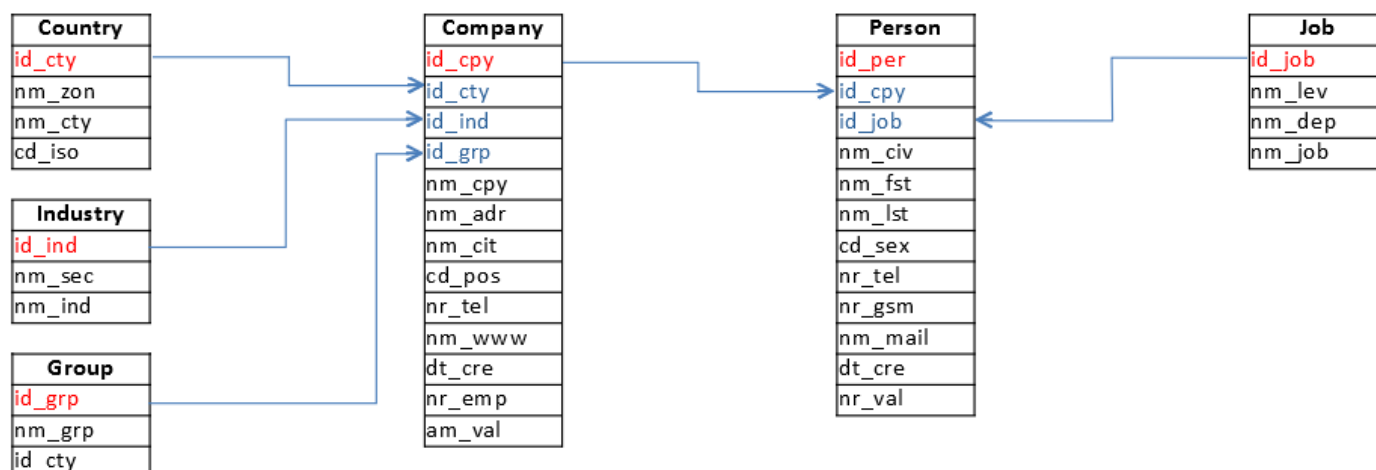
On se concentrera sur:

- la création d'un fichier base de données unique, comme spécifié ultérieurement
- la recherche dichotomique pour les clés primaires de person et company
- la création d'un index sur la clé étrangère id_cpy de la table person
- la création d'un index sur le nom de famille nm_1st de la table person
- la génération de rapports, sous forme de fichiers textes formatés
- l'affichage en mode console des données via des listes chaînées

On ne demande pas:

- de modifier les données, mode «read-only»
- de modifier la structure et de rééquilibrer les arbres
- une interactivité poussée, les écrans en mode console seront propres et bien alignés mais simples

Schéma indicatif de la base de données:



Signification des champs :

Country : Codes pays		Industry : Codes des industries	
id_cty	Clé primaire numérique	id_ind	Clé primaire numérique
nm_zon	Nom de la zone géographique	nm_sec	Nom du secteur d'activité
nm_cty	Nom du pays	nm_ind	Nom de l'industrie
cd_iso	Code iso du pays	Job : Codes des fonctions	
Group : Codes des groupes de compagnies		id_job	Clé primaire numérique
id_grp	Clé primaire numérique	nm_lev	Nom du niveau de la fonction
nm_grp	Nom du groupe	nm_dep	Nom du département de la fonction
id_cty	Clé étrangère référençant un pays	nm_job	Nom de la fonction

Company : Compagnies clientes		Person : Personnes clientes	
id_cpy	Clé primaire numérique	id_per	Clé primaire numérique
id_cty	Clé étrangère référençant un pays	id_cpy	Clé étrangère référençant une compagnie
id_ind	Clé étrangère référençant une industrie	id_job	Clé étrangère référençant un job
id_grp	Clé étrangère référençant un groupe	nm_civ	Civilité
nm_cpy	Nom de la compagnie	nm_fst	Prénom
nm_adr	Adresse de la compagnie	nm_lst	Nom de famille
nm_cit	Ville ou elle est localisée	cd_sex	Indicateur de genre
cd_pos	Code postal de cette ville	nr_tel	Numéro de téléphone
nr_tel	Numéro de téléphone	nr_gsm	Numéro de gsm
nm_www	Adresse du site	nm_mail	Adresse mail professionnelle
dt_cre	Date de création dans la database	dt_cre	Date de création dans la database
nr_emp	Nombre d'employés	nr_val	Nombre d'actions détenues
am_val	Valeur d'une action		

Les données sont généralement du type chaînes de caractères ou entiers. Am_val est un float, les dt_ sont des dates.

Les fichiers .csv fournis permettront de se faire une idée des valeurs et du nombre de records.

Rapports

On demande un rapport de détail et un rapport agrégé, par exemple :

Rapports de détail :

- Liste de toutes les personnes travaillant pour chaque compagnie d'un groupe
- Liste de toutes les compagnies appartenant à un groupe, par pays

Rapports agrégés :

- Nombre de personnes par groupe (somme des nr_emp) et nombre de personnes connues dans la DB (comptage des records personnes liés au groupe)
- Valeur totale des actions détenues par les employés pour un groupe donné, éclaté par par genre et niveau de fonction.

Exemples de requêtes pour les rapports:

Détail	Agrégé
<pre>SELECT c.nm_sec, a.nm_cpy, a.nm_adr, a.nm_cty FROM company a, country b, industry c WHERE b.id_cty = a.id_cty AND c.id_ind = a.id_ind AND b.nm_zon = "Benelux" AND c.nm_sec = "Pharma" ORDER BY c.nm_sec, a.nm_cpy ;</pre>	<pre>SELECT c.nm_zon, d.nm_lev, COUNT(*) FROM company a, person b, country c, job d WHERE b.id_cpy = a.id_cpy AND c.id_cty = a.id_cty AND d.id_job = b.id_job AND d.nm_dep = "Sales" GROUP BY c.nm_zon, d.nm_lev ;</pre>

Structure physique de base de données

La base de données se composera des 9 zones suivantes :

1	Header	Contient les métadonnées	
2	Country	Codes : Accès direct	100 rec
3	Job	Codes : Accès direct	200 rec
4	Industry	Codes : Accès direct	100 rec
5	Group	Codes : Accès direct	3.000 rec
6	Company	Records : Accès dichotomique ou direct	100.000 rec
7	Person	Records : Accès dichotomique ou indexé ou direct	500.000 rec
8	Index sur person.id_cpy	Accès indexé à person	500.000 rec
9	Index sur person.nm_lst	Accès indexé à person	500.000 rec

Phase de préparation du dossier :

Le projet Agenda ainsi que les exercices proposés dans le livre de référence (CH_DICHO.c p 70 Ch2, REPTTEL.C p115 Ch4 et REPTTEL2.C p149 Ch5) constituent la base du projet, les modules suivants seront indispensables :

- une gestion de menu
- une routine de tris
- une gestion de liste doublement chaînée
- une gestion d'arbre binaire équilibré en profondeur
- un module d'importation de fichier .csv, via strtok(), exemple fourni

Analyser le projet en groupe et au cours pour bien en comprendre la portée.

Créer les premières versions suivantes du projet, le nom du projet contiendra vos initiales et un numéro de version.

- Dossier_EB_V0 : Sur base du projet Agenda, mise en place du menu et de l'environnement
- Dossier_EB_V1 : Création de la base de données
- Dossier_EB_V2 : Importation des fichiers .csv fournis
- ...

Ces versions du dossier sont évidemment compilables sans erreurs.

Documents à rendre pour le 11/01/2021 au plus tard (attention il y a un autre dossier)

- Le projet CodeBlocks complet, y compris la structure de répertoires adaptée aux fichiers utilisés
- Les fichiers de test utilisés, pas les résultats des tests, ceux-ci devront se générer par vos routines de test
- Un rapport de 5-6 pages au format .pdf reprenant:
 - La description des fonctions nécessaires à la réalisation du travail
 - La structure globale de votre fichier database
 - La description des tests effectués
 - Les problèmes rencontrés
- Deux ou trois versions intermédiaires du projet me seront envoyées.

Critères d'évaluation

Voir la grille critériée discutée au premier cours.