

LES BANQUES DE DONNEES ET LEUR SYSTEME
DE GESTION

CANQ
TR
AD-6
109



Note

Ce cahier sur les concepts relatifs aux banques de données est un document particulièrement technique pouvant être difficilement compris par les non-initiés. Il vise particulièrement les banques à accès multiples utilisées à la fois pour l'information et la gestion.

Il s'agit d'une recherche assez poussée dans un domaine nouveau où plusieurs théories visent les mêmes objectifs par des voies différentes. Les auteurs s'en sont tenus aux notions les mieux assurées, selon une approche aussi objective que possible, en synthétisant différents articles sur le sujet.

On y trouve, en particulier, les points primordiaux à analyser dans l'évaluation des avantages, des inconvénients, des contraintes et des moyens pour assurer qu'un tel système remplisse son rôle et fournisse des bénéfices réels.

Le directeur des Systèmes de gestion

Raymond Désy, ing.

RD/nbd

Québec, le 4 novembre 1975.

466916

MINISTÈRE DES TRANSPORTS
CENTRE DE DOCUMENTATION
700, BOUL. RENÉ-LÉVESQUE EST.
21^e ÉTAGE
QUÉBEC (QUÉBEC) - CANADA
G1R 5H1

LES BANQUES DE DONNEES ET LEUR SYSTEME DE GESTION

Pierre Beaudoin
René Bertrand
Pierre Prémont

Septembre 1975

CHWQ
TR
AD
109

TABLE DES MATIERES

	<u>Page</u>
TABLE DES MATIERES	
FIGURES	
1. DESCRIPTION SOMMAIRE	1
2. LA BANQUE DE DONNEES: ELEMENT FONDAMENTAL D'UN SYSTEME D'INFORMATION POUR LA GESTION	2
2.1 La définition d'un système d'information pour la gestion	2
2.2 Le concept de système	4
2.3 Le concept d'information	4
2.4 Le concept de gestion	5
2.5 Le rôle de la banque de données	5
3. LE SYSTEME DE GESTION DE LA BANQUE DE DONNEES: DEFINITION ET CARACTERISTIQUES	7
3.1 La définition d'une banque de données	7
3.2 La définition du système de gestion de la banque de données	7
3.3 Les faiblesses de l'approche traditionnelle	8
3.4 Les objectifs du système de gestion de la banque de données	9
3.5 Les recommandations et les spécifications de CODASYL	11
3.6 Les avantages et les inconvénients du système de gestion de la banque de données	14

	<u>Page</u>
4. L'ORGANISATION STRUCTURELLE DE LA BANQUE DE DONNEES	17
4.1 L'organisation traditionnelle du fichier	17
4.2 L'organisation des banques de données	18
4.3 La classification des structures logiques de données	22
4.4 La structure physique des données	25
4.5 Les méthodes d'accès traditionnelles	28
4.6 Les méthodes d'accès de la banque de données	31
4.7 La recommandation de CODASYL	34
5. LE CONTRÔLE DE LA BANQUE DE DONNEES PAR LE DICTIONNAIRE DES DONNEES	35
5.1 La définition du dictionnaire des données	35
5.2 Les fonctions du dictionnaire des données	35
5.3 Les attributs des données	36
5.4 Les avantages du dictionnaire des données	36
6. POUR OBTENIR L'INFORMATION: LE GENERATEUR DE RAPPORTS	38
6.1 La définition du générateur de rapports	38
6.2 Les objectifs du générateur de rapports	38
6.3 Les fonctions du générateur de rapports	38
7. LE RÔLE DE L'ADMINISTRATEUR DES DONNEES	40
7.1 Les responsabilités de l'administrateur des données	40
7.2 Les outils de l'administrateur des données	41
7.3 Les difficultés d'implantation d'un système de gestion d'une banque de données	43
7.4 La compétence de l'administrateur des données	44

	<u>Page</u>
8. LES "PACKAGES" DISPONIBLES	45
8.1 L'utilisation d'un "package" de gestion de données	45
8.2 IMS-VS	45
8.3 Total	47
8.4 Adabas	48
8.5 System 2000	51
8.6 Le choix d'un "package"	52
9. POUR ARRIVER À L'IMPLANTATION D'UNE BANQUE DE DONNEES	53
GLOSSAIRE	57
BIBLIOGRAPHIE	64

FIGURES

	<u>Page</u>
1. Concept de système	3
2. Structure de données hiérarchique	19
3. Structure de données en un seul niveau	19
4. Configuration d'un dossier logique et d'un dossier physique	20
5. Sous-schéma de la figure 2	21
6. Sous-schéma de la figure 2	21
7. Schéma d'une banque de données	22
8. Description de certaines définitions	23
9. Termes utilisés pour décrire une structure en arbre	24
10. Structure de données en réseau	25
11. Structure de données en association	25
12. Structure de relation en arbre	27
13. Structure de relation en anneau	27
14. Exemple de liste filée	32
15. Exemple de liste inversée	33

1. DESCRIPTION SOMMAIRE

Les gestionnaires pensant de plus en plus en termes de système d'information, les informaticiens ont modifié leur approche de la conception de systèmes mécanisés pour que ces derniers produisent de l'information au lieu de faire un simple traitement de données. Cette nouvelle approche a donné naissance aux systèmes de gestion d'une banque de données.

Cette nouvelle orientation des systèmes mécanisés implique de nombreux concepts et un environnement différent : les concepts de la banque de données et l'environnement des systèmes de gestion de la banque de données.

Afin d'aider à l'orientation de l'évolution des systèmes actuels de l'organisation, le présent document se veut une référence pour les gestionnaires et les informaticiens dans leurs relations actuelles ou futures avec les banques de données; sans être exhaustif, il en couvre les différents aspects.

2. LA BANQUE DE DONNEES : ELEMENT FONDAMENTAL D'UN SYSTEME D'INFORMATION POUR LA GESTION

2.1 La définition d'un système d'information pour la gestion

Les gestionnaires ont oeuvré depuis longtemps avec les ressources humaines, matérielles et monétaires pour réaliser les objectifs de leur organisation. Ils oeuvraient aussi, plus ou moins consciemment, avec une autre ressource : l'information.

L'importance de l'information s'est accrue considérablement au cours des dernières années pour devenir une préoccupation quotidienne des gestionnaires. Cet aspect du "management" est vital puisque l'information est intimement reliée au processus de prise de décision; il peut y avoir une très forte relation entre la qualité de l'information et la qualité de la prise de décision.

Cet aspect du management, intégré à la science des systèmes, est connu sous le sigle de M.I.S. ("Management Information System") traduit par l'expression "Système d'information pour la gestion".

La définition d'un M.I.S. a soulevé une polémique entre les spécialistes de la question et on peut retenir la définition de Walter Kennevan :

"Un M.I.S. est une méthode organisée pour fournir de l'information passée, présente et projetée reliée aux opérations internes et à l'environnement. Il supporte les fonctions de planification, de contrôle et d'opération d'une organisation en fournissant une information uniforme au moment approprié pour assister le décideur".(8)

Cette définition, quoiqu'idéalisée, intègre l'objectif de base de tout système d'information pour la gestion qui est de "déterminer et de fournir, le plus efficacement, efficacement et économiquement possible, ce que le gestionnaire a besoin de connaître pour faciliter la réalisation des objectifs de l'organisation".(4)

Pour une meilleure compréhension de ce qu'est un système d'information pour la gestion, il s'avère pertinent de préciser les concepts impliqués : système, information et gestion.

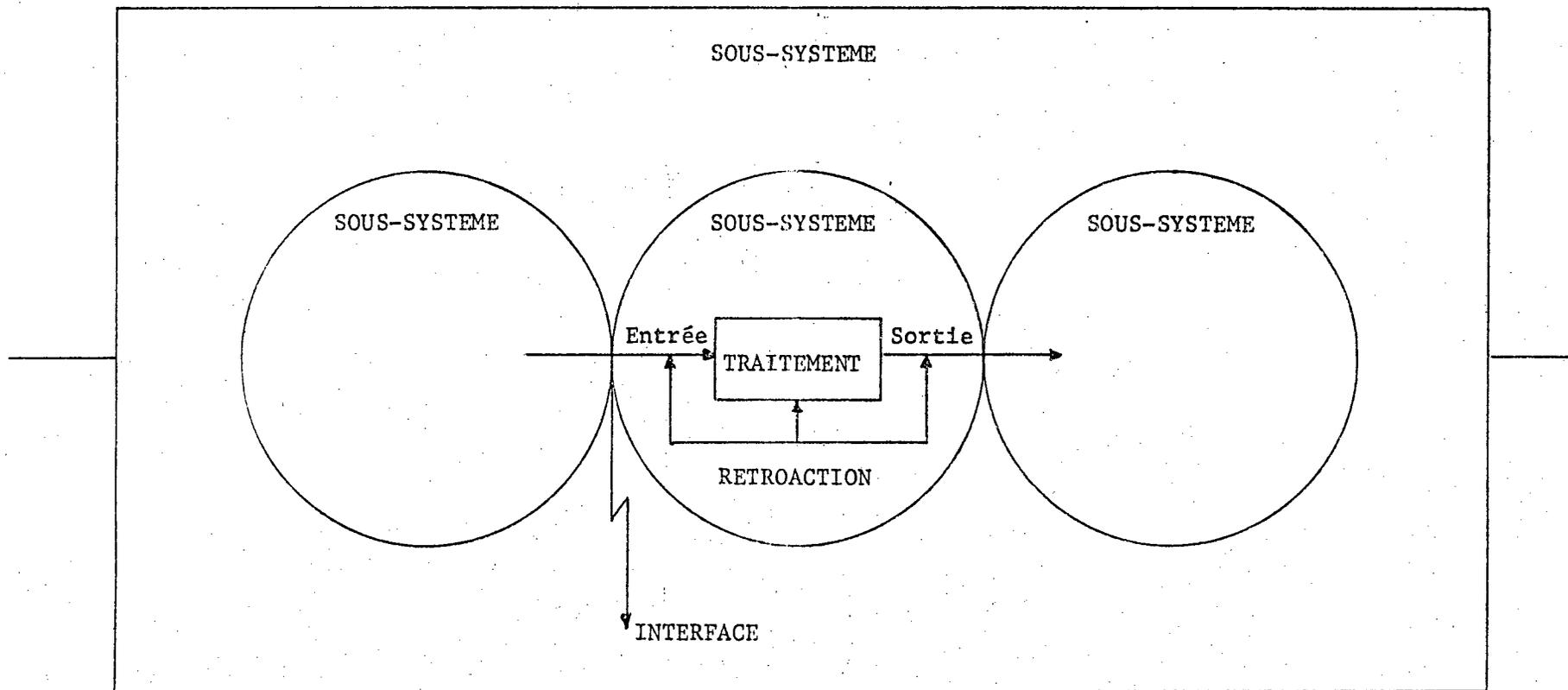


Figure 1 - Concept de système

2.2 Le concept de système (voir la figure 1)

Quoiqu'il y ait une divergence des points de vue sur la définition d'un système, on peut retenir celui de Gordon B. Davis :

"Un système est un ensemble d'éléments en interaction qui agissent pour atteindre un objectif commun. Les frontières définissent et délimitent le système. Le système est à l'intérieur des frontières; l'environnement est à l'extérieur. Chaque système est composé de sous-systèmes qui sont eux-mêmes composés d'autres sous-systèmes, chacun étant délimité par ses frontières. Les interrelations et les interactions entre les sous-systèmes génèrent des interfaces qui apparaissent aux frontières et qui prennent la forme d'entrées et de sorties".(7)

Pour régulariser et contrôler le système, il est nécessaire d'y intégrer des mécanismes de rétroaction; ceux-ci compareront les résultats actuels avec les résultats désirés pour ajuster le traitement ou les entrées au système afin que les sorties soient le plus près possible des résultats désirés.

L'identification d'un système est reliée à ses frontières qui sont définies par les personnes concernées par le système.

2.3 Le concept d'information

On ne peut parler du concept d'information sans préciser le concept de données.

Les données sont des faits bruts représentés par des symboles (chiffre, lettre, etc.); suite à un traitement et placées dans un contexte significatif, elles permettent de faire des inférences. Ces inférences sont reliées, entre autres, à la mesure et à l'identification de personnes, d'évènements et d'objets qui constituent les entités.

L'objectif de base de cueillir et de traiter des données est de produire de l'information. Cependant toutes les données traitées ne génèrent pas nécessairement de l'information appropriée et opportune.

L'information est donc "un ensemble de données qui ont été traitées dans une forme qui est significative au récepteur et qui est d'une valeur réelle ou perçue dans les décisions courantes ou à venir".(7)

L'information est l'accroissement dans la connaissance obtenue par le récepteur en réunissant les données appropriées aux variables d'un problème; c'est l'agrégation ou le traitement de données pour procurer la connaissance.

L'information est l'ingrédient-clé dans le processus de prise de décision de la majorité des individus mais elle n'est pas cependant la garantie d'une bonne prise de décision. D'autres facteurs influencent la prise de décision; ils peuvent être résumés comme suit : (24)

Le fait que le fournisseur de données remplisse sa tâche de fournir des données précises, complètes et à temps n'implique pas que le récepteur transforme ces données reçues en information. Il y a plusieurs raisons à cela. Une de celles-ci est que la transformation de données en information implique la notion de pertinence laquelle est extrêmement difficile à déterminer à l'avance dans la recherche de l'information. De plus, la transformation de données en information implique aussi l'habileté du récepteur à comprendre intellectuellement les données fournies.

La transformation de données en information est aussi fonction de l'environnement physique par lequel les données sont communiquées et de l'ensemble des expériences acquises par le récepteur.

2.4 Le concept de gestion

La gestion est l'ensemble des fonctions que doit assumer une organisation pour réaliser ses objectifs avec l'aide des ressources disponibles. Ces fonctions sont la planification, l'organisation et le contrôle.

Ces fonctions sont accomplies par la prise de décision qui doit être appuyée par de l'information lorsqu'elle est requise et pertinente.

2.5 Le rôle de la banque de données

On reconnaît généralement qu'un M.I.S. ne requiert pas nécessairement la présence d'un ordinateur : un système manuel peut rencontrer l'objectif de base d'un M.I.S.

Cependant la quantité et la variété des données à traiter pour fournir de l'information exige très souvent l'aide de l'ordinateur. L'utilisation de l'ordinateur n'implique pas que toutes les données nécessaires aux décisionnaires y seront conservées et traitées; l'ordinateur a ses limites et ne doit conserver et traiter que les données appropriées aux besoins de l'organisation.

Généralement, les gestionnaires désirent que leur système d'information rencontre quatre objectifs :(16)

- a) il doit avoir la souplesse nécessaire pour répondre aux changements rapidement et efficacement et plus particulièrement à ceux des besoins en information et de la technologie;
- b) il doit être capable de satisfaire les besoins d'information standards et prévisibles ainsi que les demandes d'information imprévisibles des gestionnaires;
- c) il doit faire le meilleur usage possible des personnes, des équipements, des télécommunications et des procédures qui y sont impliqués;
- d) il doit fournir une information précise, à temps, pertinente et complète.

L'efficacité d'un M.I.S. sera basée principalement sur les éléments interreliés suivants :(4)

- a) l'existence d'une banque de données;
- b) l'intégration des diverses applications fonctionnelles adjacentes;
- c) l'accessibilité aux données pour la production de rapports de gestion;
- d) la possibilité de simulation;
- e) l'accès en temps réel de la banque de données.

La banque de données est l'élément fondamental autour duquel gravite les autres éléments nécessaires au bon fonctionnement d'un M.I.S. Son rôle sera donc de satisfaire la plupart des besoins en information de l'organisation.

3. LE SYSTEME DE GESTION DE LA BANQUE DE DONNEES : DEFINITION ET CARACTERISTIQUES

3.1 La définition d'une banque de données

On définit généralement une banque de données comme suit :

C'est la collection d'un ou plusieurs ensembles de données reliés logiquement pour satisfaire les besoins en information de l'organisation.

Une banque de données implique donc :

- a) une cueillette des données;
- b) un regroupement de données selon certains critères;
- c) un lien logique entre les divers groupes ou ensembles de données;
- d) l'identification des besoins en information de l'organisation.

La banque de données concerne donc la définition des données pertinentes à l'organisation indépendamment du support de celles-ci; en soi, la banque de données est intangible.

3.2 La définition du système de gestion de la banque de données

Dans le cadre du support de la banque de données par l'ordinateur, il y aura la réunion des éléments suivants qui constitueront le système de gestion de la banque de données :

- a) des programmes pour recueillir, relier logiquement et traiter les données;
- b) de la mémoire pour conserver les données;
- c) un interface utilisateur-ordinateur pour l'accès aux données.

On notera qu'une collection de fichiers ne constitue pas nécessairement un système de gestion d'une banque de données. Il doit y avoir des mécanismes qui permettent de relier logiquement les données entre elles.

3.3 Les faiblesses de l'approche traditionnelle

L'approche de la banque de données fait suite à de nombreuses déficiences des systèmes conçus selon l'approche traditionnelle. Les principales faiblesses observées sont :(3)

a) la redondance des données

Les fichiers ont été traditionnellement conçus pour supporter des applications individuelles telles que la paie, le personnel, les stocks, etc. Dans la plupart des cas, ces fichiers contiennent des données redondantes. Ces données sont souvent mises à jour pour certains fichiers tandis qu'elles ne le sont pas, ou le sont partiellement, pour les autres. De plus, cette mise-à-jour nécessite un entretien redondant des programmes.

b) la dépendance des données et des programmes

Lorsque la même donnée réside dans différents fichiers, elle existe normalement dans différents formats. Cette variance dans le format de données communes signifie que les programmes sont taillés pour des formats spécifiques et même pour des unités physiques de conservation spécifiques.

Lorsque les dossiers des fichiers sont modifiés, tous les programmes utilisant ces fichiers doivent normalement être changés, ce qui entraîne un état perpétuel de changement aux programmes qui augmente appréciablement le coût du traitement des données.

c) l'accès par l'identificateur primaire seulement

L'identificateur primaire d'un dossier est l'identificateur du dossier ou le champ de contrôle.

Plusieurs applications exigent d'accéder à seulement une partie des dossiers à l'intérieur du fichier ou à une partie des dossiers ayant un contenu commun dans un ou plusieurs champs. Ce dernier aspect est connu comme étant l'accès au contenu.

Avec les fichiers conventionnels, il est nécessaire de lire tous les dossiers dans le fichier même si les dossiers avec le contenu requis représentent un faible pourcentage des dossiers dans le fichier.

d) le temps requis pour produire des rapports de gestion

Dans la majorité des cas, les demandes d'information nécessaires à la prise de décision ne sont pas satisfaites à temps à cause du délai de programmation de ces rapports. Comme les demandes

des gestionnaires sont très souvent imprévisibles et visent à répondre à des questions bien spécifiques, celles-ci sont souvent retournées sans réponse ou s'avèrent impossibles à satisfaire dans les délais requis.

L'approche de la banque de données vise à éliminer ces faiblesses afin que les systèmes d'information répondent aux attentes des gestionnaires. Le système de gestion de la banque de données ne peut cependant pas répondre aux demandes d'information si les données requises ne sont pas supportées. En contrepartie, lors de la conception du système, il faudra faire des compromis dans le choix des données à supporter, la principale contrainte étant le coût.

3.4 Les objectifs du système de gestion de la banque de données

En novembre 1970, le groupe GUIDE/SHARE, représentant une quarantaine de compagnies, d'universités et d'agences gouvernementales, publiait son rapport sur les standards de développement d'un système de gestion de banque de données dans lequel on retrouve les objectifs d'un tel système. Ces objectifs, qui sont d'ailleurs sensiblement les mêmes que ceux identifiés par le groupe de travail sur les banques de données de CODASYL, sont : (10)

a) l'indépendance des données

Le coût et le temps associés pour effectuer des changements insignifiants dans le traitement des données a eu de sérieux impacts sur l'habileté des systèmes d'information à fournir un service efficace et efficient aux utilisateurs. Dans le traitement des données traditionnel, les programmes et les données sont dépendants.

Si des segments de données sont modifiés, ajoutés ou enlevés de dossiers, tous les programmes d'application utilisant ce fichier doivent normalement être changés même si les segments affectés ne sont pas utilisés dans plusieurs programmes.

L'indépendance des données peut être réalisée en séparant les spécifications des données et des fichiers des spécifications des programmes; cette séparation s'effectuera par les programmes qui demandent les éléments d'information par le nom, sans égard au nombre de segments dans le dossier ou à l'ordre des éléments d'information dans les segments.

Avec l'indépendance des données, il est possible de changer les segments sans recompiler les programmes à moins qu'ils soient logiquement affectés par le changement.

b) la non-redondance des données

Dans une banque de données un segment de données peut participer dans plus d'une structure logique de données. En permettant à un segment de données d'exister une seule fois et de participer dans plusieurs structures logiques de données, le système de gestion de la banque de données permet la réduction ou l'élimination de la redondance des données.

c) la relation entre les données

Une entité donnée peut avoir certains attributs qui peuvent être nécessaires de relier de façon complexe. L'entité peut être une personne. Les attributs peuvent être le nom, l'adresse, le salaire, l'expérience, l'éducation.

Dans l'environnement d'une banque de données, il doit être possible avec la même banque de données non-redondantes de sélectionner à la fois sur les entités et à la fois sur les attributs.

Les relations de données dans cette sorte d'environnement complexe sont habituellement indiquées par des pointeurs, lesquels forment des chaînes permettant des méthodes d'accès multiples à tout dossier dans la banque de données.

d) l'intégrité des données

Avec une variété d'utilisateurs ayant accès au même système de gestion de la banque de données, il est impératif que le contrôle soit maintenu pour permettre d'assurer l'intégrité des données et celle de leurs relations aux autres données.

Pour réduire les dangers associés à des chaînes brisées, des pointeurs à double sens doivent être utilisés pour que les chaînes puissent être reconstituées dans les deux directions.

Pour prévenir des changements malencontreux aux dossiers, il est possible de restreindre l'accès pour des applications particulières à la lecture seulement, à la lecture et à la mise-à-jour seulement, à l'addition et à l'annulation et à l'entretien des structures de pointeurs.

e) l'accessibilité aux données

L'accessibilité aux données doit être contrôlée et la difficulté d'avoir accès aux données devrait être une fonction de l'autorisation donnée à chaque utilisateur.

f) la performance

L'utilité d'un système d'information est fonction du temps nécessaire pour répondre aux demandes d'information des utilisateurs.

Une transaction peut affecter plusieurs dossiers, ce qui exige plusieurs accès. L'application peut exiger que l'accès soit fait à plusieurs niveaux dans la hiérarchie des données. Ces considérations font de l'accès rapide une caractéristique importante des systèmes de gestion d'une banque de données.

L'accès rapide est généralement possible en randomisant sur l'identificateur primaire pour accéder aux segments maîtres et en utilisant des pointeurs pour accéder aux segments de détail. Certains systèmes utilisent des index pour localiser les segments maîtres.

g) la compatibilité

L'investissement dans le traitement des données prend généralement la forme de programmes, de procédures et de données. Avec le temps, cet investissement tend à augmenter considérablement. Cet investissement doit être protégé lorsque des changements dans la technologie surviennent tant au niveau des programmes que des équipements. L'indépendance des unités physiques de conservation est une forme de compatibilité.

La réalisation de ces objectifs permettra au système de gestion de la banque de données de jouer son rôle à l'intérieur des systèmes d'information de l'organisation et de satisfaire pleinement les besoins des différents utilisateurs.

3.5 Les recommandations et les spécifications de CODASYL

Afin de guider les concepteurs de systèmes de gestion d'une banque de données, le groupe de travail sur les banques de données de CODASYL a présenté un premier rapport en avril 1971. Ses recommandations sont les suivantes :(3)

a) permettre aux données d'être structurées de la manière la plus convenable à chaque application, indépendamment du fait que

certaines ou toutes ces données puissent être utilisées par d'autres applications (une telle flexibilité devant être réalisée sans exiger de redondance des données);

- b) permettre à plus d'un utilisateur d'accéder et de mettre à jour concurremment les données de la banque de données;
- c) fournir et permettre l'usage d'une variété de stratégies de recherches sur la banque de données ou sur des parties de la banque de données;
- d) fournir la protection de la banque de données contre les accès non autorisés aux données et contre les interactions malencontreuses de programmes;
- e) fournir la facilité de centraliser le contrôle de l'enregistrement des données;
- f) rendre les programmes indépendants des unités physiques de conservation des données;
- g) permettre la déclaration d'une variété de structures de données allant de celles n'ayant aucune relation entre les éléments d'une structure jusqu'aux structures de réseau;
- h) permettre à l'utilisateur d'interagir avec les données sans être concerné avec les mécanismes de maintenance des associations logiques;
- i) permettre aux programmes d'être aussi indépendants que possible des données de la banque de données et des données inhérentes à un programme;
- j) fournir une description de la banque de données qui n'est pas restreinte à un langage de programmation particulier;
- k) fournir une architecture qui permette la description de la banque de données pour rendre possible l'interface avec de multiples langages de programmation.

Dans ce même rapport, CODASYL a spécifié six éléments qui devraient se retrouver à l'intérieur des systèmes de gestion d'une banque de données; ce sont :(10)

- a) le langage descriptif des données

Le langage descriptif des données ("Data Description Language": D.D.L.) est utilisé pour décrire la banque de données ou une

partie de celle-ci. Ces descriptions sont faites en termes de noms et de caractéristiques des éléments d'information, des segments de données, des dossiers logiques de la banque de données et des ensembles de données.

Le concept traditionnel de fichier n'est pas inclus dans les spécifications de CODASYL, ce dernier préférant le terme "ensemble de données" parce que dans plusieurs cas des dossiers de plusieurs types seront entremêlés dans une aire de mémoire.

b) le langage de manipulation des données

Le langage de manipulation des données ("Data Manipulation Language" : DML) est utilisé par le programmeur pour transférer les données entre son programme et la banque de données. Toutes les références à la banque de données pour accéder aux données, pour ajouter de nouvelles données, pour modifier les données existantes ou les relations entre les données et pour annuler les données existantes ou les relations entre les données doivent être disponibles dans le DML.

Le DML n'est pas un langage complet en lui-même. Il relie le langage traditionnel de programmation avec les exigences procédurales requises dans le programme d'application. Le langage traditionnel est alors le langage utilisé pour manipuler les données dans la mémoire centrale de l'ordinateur; le DML est le langage utilisé pour faire l'interface avec la banque de données.

c) les routines d'utilité ou de service

Ce sont les routines requises pour supporter une banque de données dans ses opérations régulières.

d) le système de recouvrement de la banque de données

Ce système inclut les facilités pour recueillir et pour conserver les différentes transactions qui ont affecté la banque de données, ainsi que les facilités pour reconstituer la banque de données à partir de ces transactions.

e) le langage pour modifier le schéma ou le sous-schéma

"Schéma" est le terme adopté par CODASYL pour représenter les instructions du DDL qui fournissent une description complète de la banque de données. Il inclut les noms et les descriptions de tous les ensembles de données, de tous les dossiers et de tous les segments de données de la banque de données.

"Sous-schéma" est le terme représentant les instructions de DDL définissant le sous-ensemble du schéma qui est d'intérêt pour un ou plusieurs programmes d'application. Il décrit seulement les ensembles de données, les dossiers et les segments de données de la banque de données connus de ces programmes.

Une modification à un schéma signifie une modification aux dossiers logiques de la banque de données.

f) le langage de contrôle des unités physiques

Ce langage contrôle l'assignation des dossiers physiques aux unités physiques de conservation compte tenu de l'espace requis et disponible en contrôlant le "buffering", le "paging" et l'"overflow".

Ces spécifications, quoique préliminaires, permettent d'envisager l'organisation générale d'un système de gestion d'une banque de données et de constater comment pourrait être réalisés les objectifs de la banque de données.

3.6 Les avantages et les inconvénients du système de gestion de la banque de données

Le développement et l'implantation d'un système de gestion d'une banque de données présentent de nombreux avantages; ce sont : (3)

- a) la possibilité d'organiser les données ou la manière la plus convenable et la plus appropriée aux fonctions interreliées de l'organisation;
- b) la description des données est contenue dans la banque de données indépendamment des fonctions de programmation, écartant ainsi les programmeurs de la gestion des données;
- c) la possibilité de fournir aux utilisateurs un interface direct avec la banque de données;
- d) l'intégration des éléments d'information pour minimiser la redondance;
- e) la possibilité d'évoluer sans une revision majeure du système;
- f) la réponse rapide aux besoins des utilisateurs;

- g) la possibilité de rencontrer les besoins changeants des utilisateurs dans le temps (les utilisateurs du système d'information ne peuvent pas anticiper tous leurs besoins ni ne peuvent garantir que les besoins actuels tels qu'établis demeureront constants);
- h) la possibilité de rencontrer des changements dans le management sans une révision majeure du système (si la banque de données est organisée sur la base des relations fonctionnelles, représentant ainsi le modèle des opérations et le flot des activités, alors un changement majeur dans les utilisateurs du système ne devrait pas créer un besoin pour des changements significatifs dans la conception générale de la banque de données);
- i) la mise-à-jour des données se fait simultanément, i.e. que lorsqu'une transaction survient, toutes les données reliées à cette transaction sont mises à jour pour refléter ce changement;
- j) les erreurs et les inconsistances dans les données sont réduites parce que la redondance est réduite;
- k) un système d'information plus vaste, mieux coordonné et plus fiable peut être fourni aux utilisateurs de l'organisation;
- l) des économies sont réalisées;
- m) l'interrogation directe des données est disponible avec une réduction concomitante dans les sorties volumineuses de rapports;
- n) la manipulation des données est réduite parce que plusieurs duplications d'entrée et de transfert de données sont éliminées;
- o) la simplification des travaux de programmation;
- p) l'efficacité du traitement est améliorée par la réduction du nombre de tris.

A ces avantages sont reliés certains inconvénients qui ne peuvent être ignorés :

- a) la conception et l'implantation d'un système de gestion d'une banque de données exigent des professionnels hautement qualifiés;

- b) l'investissement initial est extrêmement élevé;
- c) les opérations de l'organisation peuvent être interrompues lors d'un mauvais fonctionnement de l'ordinateur ou lors d'erreurs de programmation;
- d) un niveau sophistiqué de programmation et d'équipements est nécessaire;
- e) un niveau élevé de mesures de protection et de copies de sécurité est requis;
- f) une multitude d'erreurs peuvent se développer à travers la banque de données suite à une seule erreur.

L'engagement d'une organisation dans un système de gestion d'une banque de données représente un investissement en ressources humaines et monétaires important. Il doit donc être évalué de la même façon que tout autre investissement.

4. L'ORGANISATION STRUCTURELLE DE LA BANQUE DE DONNEES

4.1 L'organisation traditionnelle des fichiers

Traditionnellement, les fichiers furent conçus pour accommoder des applications individuelles. On s'est rarement attardé à standardiser le format des fichiers, à contrôler les cycles de mise à jour ou à minimiser les éléments d'information redondants dans les fichiers.

Les applications traditionnelles étaient caractérisées comme suit: (3)

- a) un ou plusieurs programmes étaient écrits pour mettre à jour le fichier maître à partir des transactions;
- b) un ou plusieurs programmes étaient nécessaires pour changer ou pour modifier le fichier maître lorsque des erreurs étaient détectées;
- c) un ou plusieurs programmes étaient requis à chaque fois qu'un rapport était désiré;
- d) un ou plusieurs programmes spéciaux étaient nécessaires pour mettre en relation deux ou plusieurs fichiers.

La structure des dossiers des fichiers était à un seul niveau avec une longueur fixe, variable ou indéfinie où les éléments d'information étaient adressés selon leur position relative dans le dossier.

Plus tard, on tenta de minimiser ou de résoudre le problème de la gestion des fichiers en standardisant l'accès aux données de ces fichiers. Les systèmes de gestion de fichiers exigeaient que le programmeur fournisse seulement le format du fichier et le format de la sortie requise. Une fois ces paramètres communiqués, un programme général lisait le fichier, en entrait les éléments d'information, complétait des opérations logiques et arithmétiques limitées et arrangeait la sortie dans le format désiré.

Par la suite, on a ajouté les facilités suivantes:

- a) le traitement de plus d'un fichier en même temps;

- b) le traitement de fichiers organisés autrement que séquentiellement;
- c) l'expansion des possibilités logiques et arithmétiques;
- d) la production de différents formats de sortie;
- e) le traitement avec des terminaux interactifs tels que les écrans cathodiques et les télétypes;
- f) la mise à jour des fichiers.

Les systèmes de gestion de fichiers ont surtout permis de séparer le format du fichier des programmes d'application accédant au fichier et de mettre à la disposition des programmeurs et des utilisateurs des programmes généraux de traitement.

4.2 L'organisation des banques de données (9)

Les buts premiers de l'organisation des données en banque de données furent d'abord:

- a) d'effectuer l'indépendance des données;
- b) de réduire la redondance des données;
- c) de relier logiquement et de rendre consistantes les données avec les aspects fonctionnels des utilisateurs de l'organisation;
- d) de réduire le coût du traitement et de conservation des données.

L'indépendance des données permettrait ainsi que les données puissent être utilisées par plus d'un programme concurremment tout en permettant des changements dans la définition des données sans un changement conséquent dans les programmes d'application.

Pour réaliser ces buts, il fallait donc que la structure logique des données puisse être envisagée comme un ensemble de segments de données reliés hiérarchiquement.

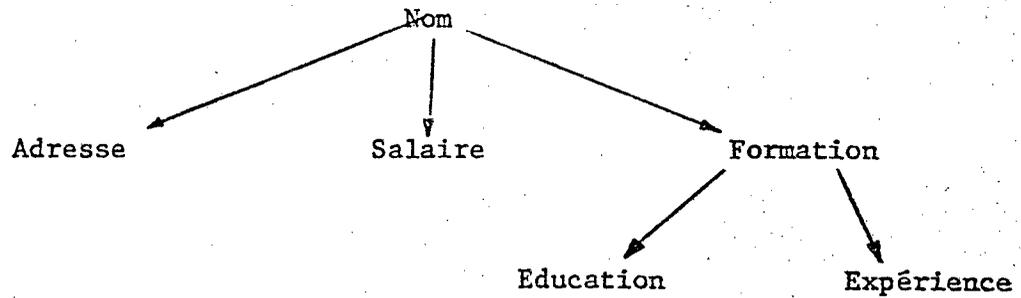


Figure 2 - Structure de données hiérarchique

L'organisation traditionnelle des fichiers (séquentielle, séquentielle indexée, aléatoire et aléatoire indexée) ne permettait pas de supporter les structures hiérarchiques de données. Les méthodes traditionnelles ne permettaient que de supporter des ensembles de dossiers à un seul niveau.

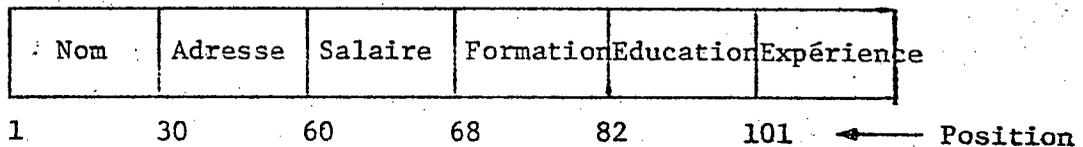
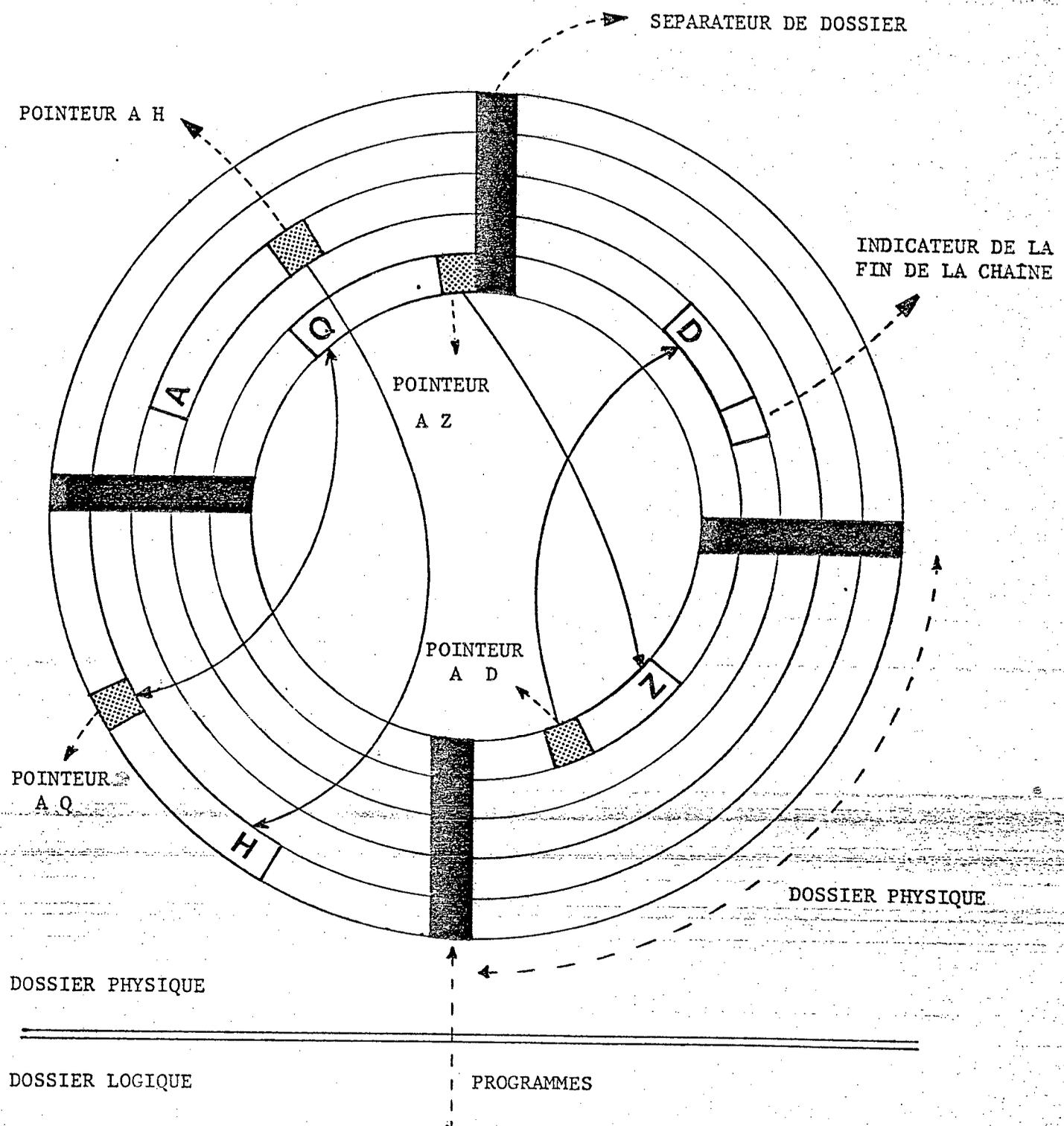


Figure 3 - Structure de données en un seul niveau

L'approche de la banque de données dans la conception de système diffère de l'approche traditionnelle parce qu'elle supporte la nature hiérarchique de la structure logique des données.

Dans la figure 2, le segment du nom avec ses cinq segments subordonnés reliés constituent un dossier logique de la banque de données. Il peut inclure des ensembles contigus de segments de données à l'intérieur des blocs physiques reliés. Dans ce cas, le dossier logique de la banque de données est représenté de façon interne comme un dossier physique de la banque de données (Voir figure 4).

Alternativement, le dossier logique de la banque de données peut être composé de segments de données de plusieurs dossiers physiques de la banque de données. La relation logique entre les segments de données, lesquels peuvent être répartis à travers la banque de données, est faite avec l'aide de pointeurs.



DOSSIER A
DOSSIER H
DOSSIER Q
DOSSIER Z
DOSSIER D

Figure 4: Configuration d'un dossier logique et d'un dossier physique (Source: 30)

Le concept de l'interrelation hiérarchique des segments rend possible l'idée fondamentale de l'indépendance des données des programmes, des méthodes d'accès et de l'organisation physique.

De plus, par le concept de la sensibilité des programmes, un programme est écrit de telle manière qu'il ne "voit" que les segments de données qui sont pertinents au traitement concerné.

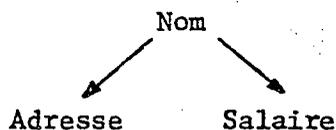


Figure 5 - Sous-schéma de la figure 2.

Le programmeur assure l'indépendance des données en spécifiant trois définitions:

- a) les segments de données à l'intérieur du dossier logique de la banque de données auxquels il désire que son programme soit sensible;
- b) la structure du dossier logique de la banque de données représenté par un ou plusieurs segments d'un ou plusieurs dossiers physiques;
- c) l'organisation de la banque de données et les méthodes d'accès qu'il utilisera.

Les segments de données reliés hiérarchiquement peuvent avoir des structures d'une variance considérable grâce aux pointeurs.

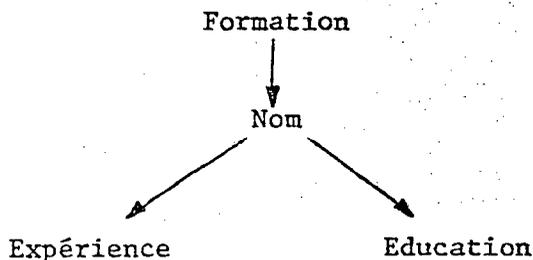


Figure 6 - Sous-schéma de la figure 2.

Les segments de données peuvent ainsi participer à plus d'une structure logique de données. En permettant à un segment de données d'assister une seule fois dans la banque de données et de participer dans plusieurs structures logiques de données, la banque de données permet de réduire ou d'éliminer la redondance des données.

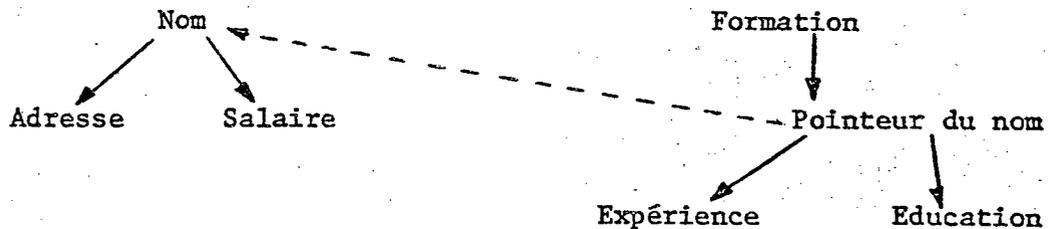


Figure 7 - Schéma d'une banque de données

4.3 La classification des structures logiques de données (9)

Les structures logiques de données peuvent être classifiées en quatre catégories :

- de niveau simple
- en arbre hiérarchique ou logique
- en réseau ou graphe
- en association

Chacune de ces structures logiques de données peut être décrite comme suit :

a) le niveau simple

La structure de niveau simple (voir figures 3 et 8) est la juxtaposition d'éléments d'information accessibles par un identificateur primaire seulement. Les méthodes traditionnelles d'organisation de fichiers furent développées pour supporter des dossiers de niveau simple.

b) l'arbre hiérarchique ou logique

Dans cette structure, il y a un segment de données supérieur sous lequel on retrouve des segments de données de niveau inférieur entre lesquels il existe un lien logique (voir figures 2 et 9). Le segment de données supérieur est appelé le "segment maître" et les segments de données inférieurs sont appelés les "segments de détail".

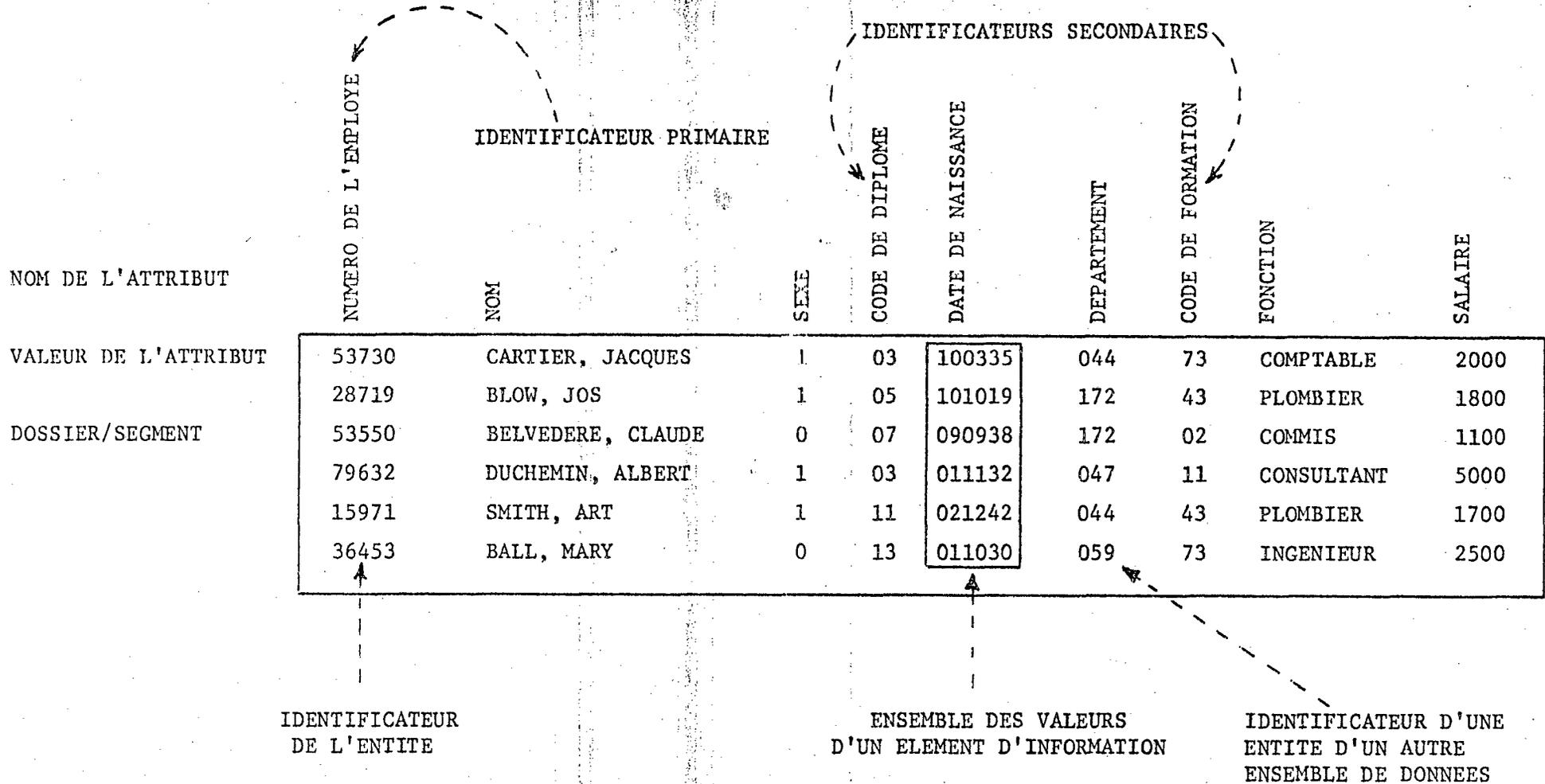
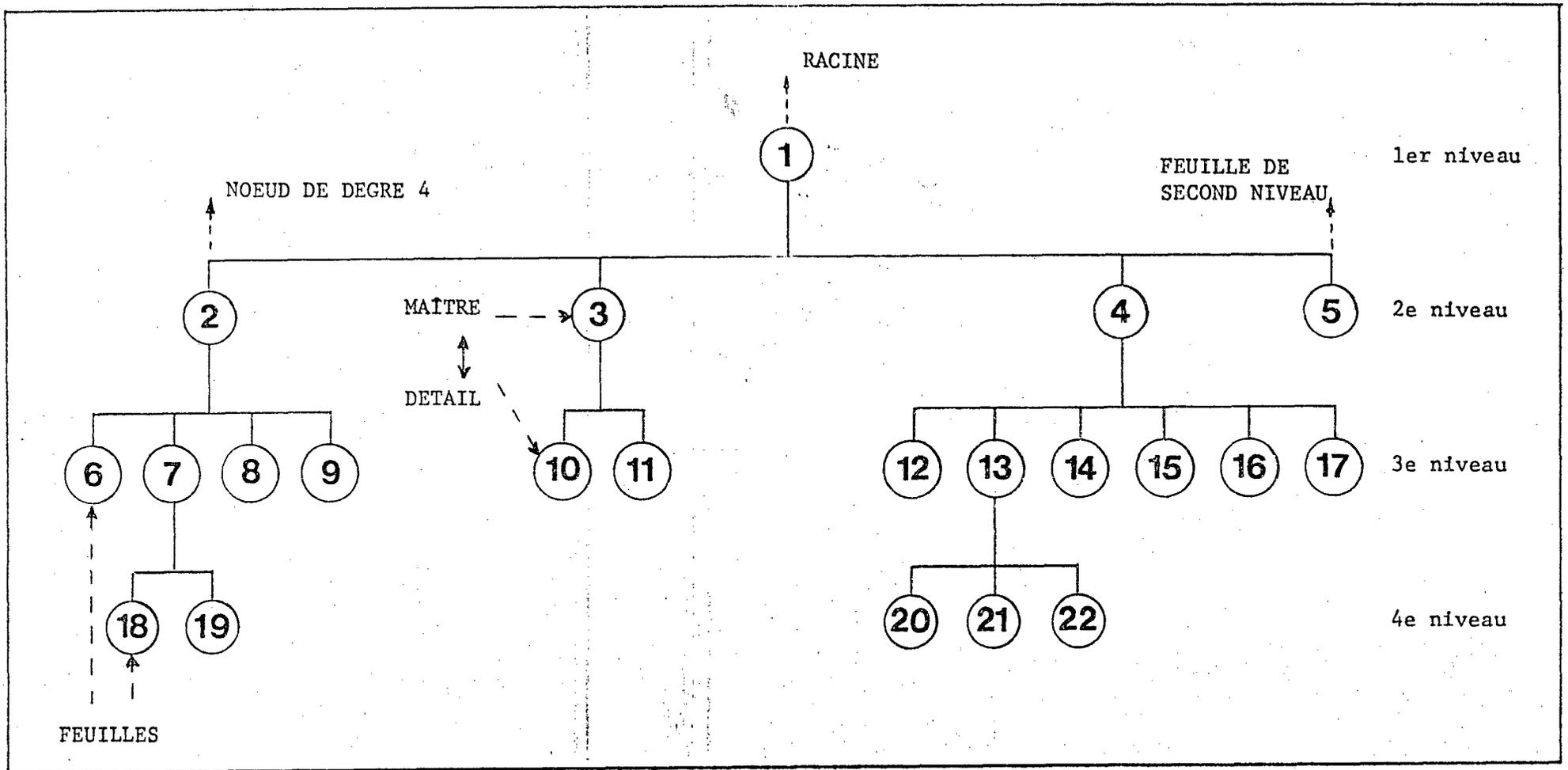


Figure 8: Description de certaines définitions (Source: 30)



- CARACTERISTIQUES DE L'ARBRE:
- hauteur de 4 (nombre de niveaux)
 - moment de 22 (nombre de noeuds)
 - poids de 16 (nombre de feuilles)
 - base de 1 (nombre de racines)

Figure 9: Termes utilisés pour décrire une structure en arbre (Source: 30)

c) le réseau en graphe

La caractéristique prédominante de cette structure de données est que le segment de détail est partagé par plus d'un segment maître (voir figure 10).

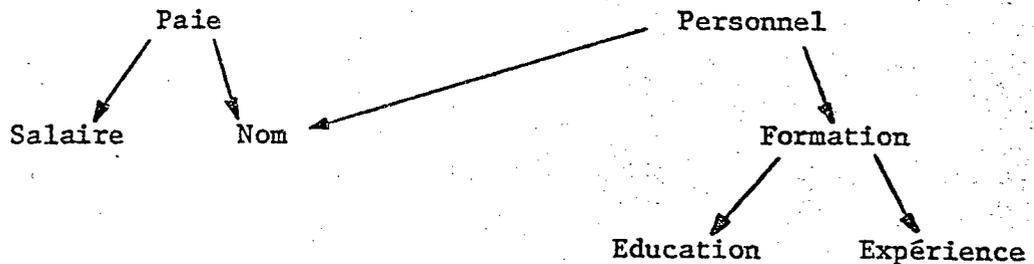


Figure 10 - Structure de données en réseau

d) l'association

Une structure de données en association est une structure dont les segments de données sont reliés de telle façon que chacun contient une valeur qui est identique. Pour une formation donnée, des pointeurs peuvent être utilisés pour pointer à tous les segments de nom de la banque de données qui sont reliés à cette formation (voir figure 11).

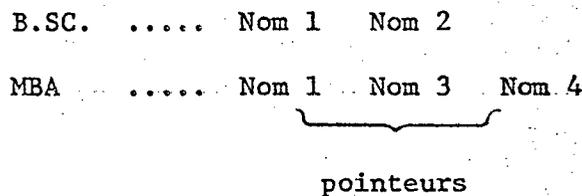


Figure 11 - Structure de données en association

4.4 La structure physique des données (9)

Deux concepts importants sont reliés aux structures physiques des données:

- les méthodes de l'enregistrement
- les techniques de mise en relation.

Chacun de ces concepts peut être décrit comme suit:

a) les méthodes d'enregistrement

Les méthodes d'enregistrement sont reliées au problème de la conservation des dossiers logiques de la banque de données dans une aire de mémoire selon les caractéristiques du média de mémoire physique et des méthodes d'accès disponibles. Il existe deux méthodes d'enregistrement:

- l'enregistrement séquentiel est utilisé lorsque les dossiers logiques de données se suivent dans l'aire de mémoire (avec l'enregistrement séquentiel, l'accès séquentiel consiste à lire le prochain dossier);
- l'enregistrement distribué est utilisé lorsque les dossiers logiques sont enregistrés à des endroits connus dans l'aire de mémoire et que des méthodes existent pour aller directement à ces endroits.

Il est aussi possible de combiner ces deux méthodes d'enregistrement pour enregistrer les dossiers logiques de la banque de données.

b) les techniques de mise en relation

Lorsque deux ou plusieurs champs de données sont reliés, la méthode la plus commune pour indiquer cette relation est de les réunir pour former un segment de données. La méthode la plus commune est d'utiliser des champs de longueur fixe, où la location des éléments d'information est connue à partir de sa position dans le segment de données ou dans le dossier (voir figure 3).

La relation entre des segments de données réfère à la méthode d'exprimer les relations maître-détail entre des segments de données dans une structure d'enregistrement des données. La méthode la plus commune est celle d'adjoindre séquentiellement les segments de données (de la même façon que les systèmes à ruban magnétique).

Dans les systèmes de mémoire à accès direct où l'enregistrement distribué est utilisé, les pointeurs sont communément utilisés pour relier les segments de données. Ces pointeurs peuvent être soit imbriqués dans des segments de données ou extraits et enregistrés dans des index.

Lorsque des pointeurs fournissent une communication hiérarchique, il y a toujours un segment maître et des segments de détail. Deux

méthodes communes pour implanter des structures de données hiérarchiques au moyen de pointeurs sont:

- la structure de relation en arbre
- la structure de relation en anneau.

Chacune de ces deux structures de relation par pointeur peut être décrite comme suit:

- la structure de relation en arbre.

Sous cette structure de relation, les pointeurs contenus dans un segment de données de niveau supérieur pointent à des segments de niveau immédiatement inférieur sans qu'il y ait de relation entre ces segments de niveau inférieur (voir figure 12).

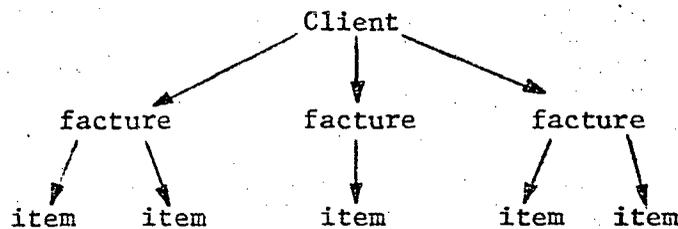


Figure 12 - Structure de relation en arbre.

- la structure de relation en anneau

Sous cette structure de relation, les pointeurs contenus dans les segments de données de niveau consécutif permettent de compléter une boucle (voir figure 13).

Avec la structure de relation en anneau, il est commun d'organiser les segments maîtres de façon aléatoire, séquentielle indexée et aléatoire indexée, les segments maîtres sont localisés en référant à un index ou en randomisant sur un identificateur primaire. Les segments de détail sont normalement accédés par des pointeurs à l'intérieur des segments maîtres.

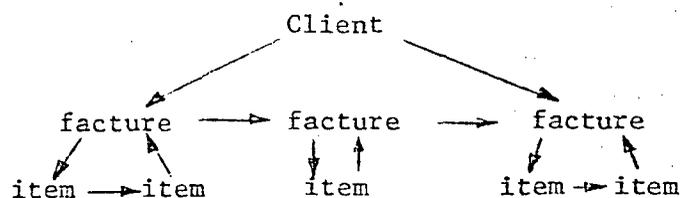


Figure 13 - Structure de relation en anneau

D'autres structures de relation peuvent être conçues par la combinaison de l'un ou l'autre des concepts exposés précédemment.

4.5 Les méthodes d'accès traditionnelles (9)

La méthode d'accès est le moyen par lequel on accède aux segments de données dans la structure d'enregistrement physique.

L'identificateur primaire est le champs de contrôle ou l'identificateur du dossier. Dans une organisation traditionnelle de fichiers, l'identificateur primaire doit être connu pour avoir accès aux données du dossier, à moins que la technique "get next" soit utilisée, ou que le fichier soit complètement lu.

Dans plusieurs situations cependant, l'accès au contenu du dossier est requis. Dans ces cas, on établit des partitions en se basant sur les valeurs contenues dans un ou plusieurs éléments d'information des dossiers. Ces valeurs sont connues comme des identificateurs secondaires.

Les méthodes d'accès les plus communes sont les suivantes:

a) la lecture complète du fichier

Cette méthode d'accès implique la lecture de tous les dossiers dans le fichier en cherchant une ou plusieurs valeurs particulières.

b) l'accès direct

On accède au dossier d'un fichier par cette méthode en calculant l'adresse du dossier à partir de son identificateur primaire.

c) l'accès "get next"

L'adresse du dossier considéré comme le suivant est contenue dans la structure des données. La forme la plus commune est l'accès séquentiel où les dossiers logiques de données sont enregistrés de façon contiguë. Une deuxième forme est possible lorsqu'une chaîne de dossiers existe lesquels sont accédés en utilisant des pointeurs imbriqués dans les dossiers.

d) l'accès indexé

Des index enregistrés dans un autre fichier fournissant les adresses des dossiers à être accédés.

En pratique, les méthodes d'accès sont souvent une combinaison de ces méthodes de base.

L'organisation d'un fichier peut être considérée comme la correspondance existant entre la structure des données et la structure d'enregistrement physique. Il y a une tendance pour accéder les dossiers par des identificateurs secondaires et de faire le traitement efficient des ensembles de données en utilisant des structures de données hiérarchiques. Cependant, il y a encore plusieurs applications qui exigent l'accès sur l'identificateur primaire seulement où les dossiers sont essentiellement de niveau simple.

Il y a quatre méthodes principales d'organisation de fichiers qui ont été développées pour supporter des dossiers de niveau simple:

a) l'organisation séquentielle

Avec l'organisation séquentielle, les dossiers sont séquentiellement distribués. Cette méthode n'est généralement pas propice pour les applications en temps réel parce qu'elle exige une recherche séquentielle et n'a pas de facilité pour une mise à jour rapide.

Cependant, si l'application en temps réel est limitée à de rares interrogations, la technique de recherche binaire peut être utilisée.

b) l'organisation séquentielle indexée

Le traitement séquentiel régulier a le désavantage que le fichier doit être recopié à chaque fois qu'il est modifié. De plus, la plupart des applications en temps réel nécessitant quelque chose de mieux que la recherche séquentielle, une alternative est d'avoir un répertoire des dossiers enregistrés. Un tel index est appelé un index d'identificateurs primaires.

Cet index relie les identificateurs primaires des dossiers à leur lieu d'enregistrement. Si les dossiers sont en séquence de la valeur de l'identificateur primaire des dossiers, l'index n'a pas besoin de contenir une entrée pour chaque dossier enregistré; il peut y avoir une entrée pour certains groupes de dossiers enregistrés. Le dossier enregistré peut alors être localisé par une courte recherche séquentielle.

L'organisation séquentielle indexée permet soit le traitement séquentiel ou le traitement aléatoire. C'est un compromis entre le traitement séquentiel et le traitement aléatoire qui permet d'exploiter les avantages de chacun. Cependant, le compromis est coûteux en termes d'espace de mémoire et de temps d'accès.

Si le fichier est volatile (s'il est sujet à un haut pourcentage d'additions et d'annulations) l'organisation séquentielle indexée est seulement possible si la réorganisation du fichier est faite fréquemment. Les additions pour un groupe sont habituellement enchaînées dans une aire de débordement et exige une recherche séquentielle. Le système se dégrade très rapidement lorsque le nombre d'additions augmente.

c) l'organisation aléatoire

L'organisation aléatoire utilise un algorithme de transformation de l'identificateur primaire pour dériver la location du dossier en mémoire. Plusieurs techniques sont connues; deux de celles-ci sont: l'adressage direct et l'adressage indirect.

- Avec l'adressage direct, pour chaque identificateur possible dans le fichier correspond une adresse relative unique. Ceci rend possible la localisation du dossier dans le fichier avec une recherche (si les mécanismes d'accès physiques sont mobiles) et une lecture. Cette méthode est seulement appropriée lorsque les dossiers sont de longueur fixe et que les identificateurs sont numériques.

L'adressage direct ne permet pas seulement le temps minimum d'accès lorsque le traitement est aléatoire, mais est aussi idéal pour un traitement séquentiel puisque les dossiers sont dans la séquence de l'identificateur primaire.

- L'adressage indirect doit être utilisé lorsque le nombre d'identificateurs primaires d'un fichier inclu un pourcentage trop élevé de dossiers inutilisés lorsque l'adressage direct est utilisé; ce dernier utiliserait alors trop d'espace. Avec l'adressage indirect, la répartition des identificateurs d'un fichier est comprimée dans une distribution plus petite d'adresser à travers un processus connu comme la randomisation.

L'adressage indirect cause inévitablement des synonymes (deux ou plusieurs dossiers dont les identificateurs randomisent à la même adresse). Ces synonymes produisent des dossiers de débordement qui doivent être chaînés, ce qui cause des problèmes lorsque le fichier est volatile.

Une façon de minimiser les synonymes est d'allouer plus d'espace au fichier qu'il est nécessaire d'avoir pour supporter tous les dossiers du fichier. Le terme "facture de groupage" signifie le pourcentage de locations assignées actuellement utilisées.

d) l'organisation aléatoire indexée

Dans une organisation aléatoire indexée, les dossiers enregistrés sont désordonnés en regard de l'identificateur primaire. Le niveau le plus bas de l'index doit fournir une entrée pour chaque dossier enregistré.

Pour le coût de cet espace supplémentaire, l'organisation offre plus de flexibilité que l'organisation séquentielle indexée. Les données peuvent être partitionnées de diverses façons et la séquence physique peut être utilisée à tout autre fin.

Cette technique est une façon efficace de supporter un fichier aléatoire dans un espace minimum. Elle se prête à un fichier volatile parce que, contrairement à l'organisation aléatoire, il n'y a pas de débordement. Ses désavantages sont l'espace de mémoire supplémentaire pour les index et le temps de recherche supplémentaire.

4.6 Les méthodes d'accès de la banque de données

Les méthodes d'accès traditionnelles peuvent être utilisées pour supporter des structures de données hiérarchiques lorsque l'accès par l'identificateur primaire est suffisant. C'est le cas pour les systèmes essentiellement opérationnels.

Cependant, on peut désirer avoir accès aux données par des identificateurs secondaires (l'accès au contenu) en plus de l'accès par l'identificateur primaire. À cette fin, on a identifié deux catégories de listes structurées: (9)

a) la liste filée

Avec la liste filée, on a un pointeur à l'intérieur du dossier pour chaque élément d'information sur lequel on désire avoir accès. Ce pointeur spécifie le prochain dossier qui contient une donnée identique. Les dossiers dans ce type de liste ne sont habituellement pas physiquement contigus (voir figure 14).

b) la liste inversée

La technique de la liste inversée exige aussi des pointeurs, mais dans ce type d'organisation, les pointeurs sont extraits des dossiers individuels pour être conservés dans un index; les pointeurs n'apparaissent donc pas explicitement dans les dossiers. Ils existent ailleurs dans le système et leur fonction est inchangée (voir figure 15).

Il est généralement désirable de rendre l'identificateur du dossier semblable à l'identificateur de la liste pour que l'accès par l'identificateur du dossier soit aussi possible. Ceci est particulièrement utile dans les systèmes avec des mises à jour en temps réel.

La méthode de la liste inversée est particulièrement intéressante pour les systèmes exigeant des accès logiques sur plusieurs identificateurs. Par un accès logique sur plusieurs identificateurs, on entend l'accès à ces dossiers qui satisfont des équations booléennes telles que:

PRINT NOM FOR TITRE CONSEILLER AND DEGRE MBA AND DISCIPLINE MIS

<u>NUMERO EMPLOYE</u>	<u>NOM</u>	<u>DIPLOME</u>
01785	CARTIER, J	44
07897	BLOW, J	50
18743	BELVEDERE, C	63
22471	DUCHEMIN, A	44
22582	SMITH, A	18
27888	BALL, M	22
38124	PROULX, R	44
39216	LITRE, V	44

<u>DIRECTOIRE</u>	
Diplôme	18 Adresse
	22 Adresse
	44 Adresse
	50 Adresse
	63 Adresse

Figure 14 - Exemple de liste filée

<u>NUMERO EMPLOYE</u>	<u>NOM</u>	<u>DIPLÔME</u>
01785	CARTIER, J	44
07897	BLOW, J	50
18743	BELVEDERE, C	63
22471	DUCHEMIN, A	44
22582	SMITH, A	18
27888	BALL, M	22
38124	PROULX, R	44
39216	LITRE, V	44

<u>DIRECTOIRE</u>	
Diplôme	18 Adresse
	22 Adresse
	44 Adresse
	Adresse
	Adresse
	Adresse
	50 Adresse
	63 Adresse

Figure 15 - Exemple de liste inversée

Les listes inversées et les listes filées sont des méthodes possibles pour organiser la même banque de données. Les mises à jour sont plus simples parce qu'on évite de maintenir deux fichiers.

Pour chaque combinaison d'un élément d'information avec une valeur donnée, il y a un pointeur dans un directoire. C'est l'adresse du premier dossier dans la chaîne.

Lorsqu'une addition est faite, le pointeur du directoire devient le pointeur du nouveau dossier qui devient le premier dossier dans la chaîne. La location physique est conservée dans le directoire. Les annulations sont simplement des codes d'annulations jusqu'à la réorganisation de la banque de données.

Le désavantage de la liste filée est l'accès logique par plusieurs identificateurs. La technique qui doit être utilisée est d'accéder tous les dossiers dans la chaîne la plus courte impliquée par la recherche afin de déterminer si le critère de recherche est rencontré.

4.7 La recommandation de CODASYL (9)

Le groupe de travail sur les banques de données de CODASYL recommande d'utiliser la liste en anneau organisée dans une structure de relation en anneau. Ces listes sont circulaires de sorte que la queue de toute liste ramène à sa tête plutôt que contenir un indicateur pour montrer que c'est le dernier dossier dans la liste.

Cette technique peut être très puissante parce qu'elle permet d'accéder à plusieurs types de dossiers, tous reliés logiquement, et de traiter tous les dossiers de l'anneau en se branchant à l'un ou l'autre de ces dossiers et ainsi accéder à d'autres dossiers qui sont reliés logiquement. De tels dossiers seraient aussi enregistrés dans une structure en anneau et permettraient à leur tour la même facilité.

La puissance et la flexibilité de cette organisation sont amoindries par un désavantage dans la recherche. Le problème de la recherche est simplement qu'il peut être long de suivre une liste chaînée sur une unité de mémoire à accès direct.

5. LE CONTROLE DE LA BANQUE DE DONNEES PAR LE DICTIONNAIRE DES DONNEES

5.1 La définition du dictionnaire des données

Le dictionnaire des données est la collection organisée des descriptions de toutes les données connues de la banque de données. Il doit inclure l'ensemble complet des descripteurs de chaque élément d'information contenu dans la banque de données.

Il contient toutes les caractéristiques pour définir un élément d'information ainsi que toute l'information nécessaire pour en avoir accès. De plus, il décrit toutes les relations définies entre les éléments d'information et les relations entre les éléments d'information et les programmes, les traitements et les entrées/sorties de la banque de données.

Il contient finalement l'information nécessaire pour décrire la structure logique et physique de la banque de données.

5.2 Les fonctions du dictionnaire des données

Les principales fonctions que doit remplir le dictionnaire des données sont:

- a) de servir d'instrument à l'administrateur des données pour contrôler la structure et les règles d'accessibilité à la banque de données;
- b) de faciliter la localisation, l'accès et la conservation des données en vue de répondre aux demandes des utilisateurs;
- c) de distinguer entre le dossier logique et le dossier physique des données;
- d) d'aider les programmeurs pour un usage efficient de la banque de données;
- e) de contenir les descripteurs des dossiers logiques et des dossiers physiques de la banque de données;
- f) de contenir l'information nécessaire pour contrôler la performance, la sécurité et l'intégrité de la banque de données.

Essentiellement, le dictionnaire des données doit être un outil de contrôle de la banque de données en fournissant la documentation appropriée aux individus concernés. Sans cet outil précieux, il serait presque impensable de pouvoir contrôler l'évolution de la banque de données et de faire en sorte que cette dernière joue efficacement son rôle.

5.3 Les attributs des données

Les attributs généraux qui doivent être supportés par le dictionnaire des données peuvent être regroupés comme suit: (20)

- a) les attributs d'identification de la donnée (nom, mnémonique, description, type, etc.);
- b) les attributs de représentation de la donnée (longueur, type de caractère, format, etc.);
- c) les attributs de localisation de la donnée (unité d'enregistrement, organisation, méthode d'accès, etc.);
- d) les attributs de relation de la donnée (interaction, interrelation etc.);
- e) les attributs de l'organisation de la donnée (nombre d'exemplaires, factures de croissance, fréquence d'utilisation, etc.);
- f) les attributs de sécurité de la donnée (mots de passe, source, responsable, etc.);
- g) les attributs d'intégrité de la donnée (validité, consistance, acceptabilité, etc.).

Ces différents attributs permettront de décrire les données de façon adéquate, ce qui permettra d'obtenir une image précise de la banque de données.

5.4 Les avantages du dictionnaire des données

L'intégration à la banque de données d'un dictionnaire des données permettra:

- a) de fournir à l'administrateur des données un contrôle des données contenues dans la banque de données;
- b) d'établir ou de renforcer les standards de programmation particulièrement dans les définitions des données;
- c) de maintenir à jour une documentation précise des données supportées par la banque de données;

- d) de faciliter les échanges d'information entre les utilisateurs et les informaticiens;
- e) de supporter plusieurs versions des structures de la banque de données pour les fins de simulation et de test.

Fondamentalement, le dictionnaire des données permettra un meilleur usage et un meilleur contrôle de la banque de données.

Les données sont une ressource importante pour l'organisation et non seulement pour le département utilisateur. Le dictionnaire des données est un outil qui permet à l'organisation de protéger ses données et d'en faire le meilleur usage possible. Il permet de mieux mesurer la façon dont les données sont utilisées, ce qui permettra d'améliorer et d'optimiser leur utilisation. Il permet de plus d'avoir une vue d'ensemble de la banque de données, ce qui permettra de connaître les impacts des changements proposées.

6. POUR OBTENIR L'INFORMATION: LE GENERATEUR DE RAPPORTS

6.1 La définition du générateur de rapports

Le générateur de rapports est un ou plusieurs programmes établis pour produire rapidement et facilement les différents rapports désirés par les utilisateurs. Ces rapports seront généralement le sommaire de données sélectionnées à partir de la banque de données selon certains critères.

Les demandes des utilisateurs doivent être facilement transmissibles à l'ordinateur par un langage hautement évolué.

6.2 Les objectifs du générateur de rapports

Le générateur de rapports doit permettre la réalisation des objectifs suivants:

- a) rendre accessible la banque de données aux différents utilisateurs pour qu'ils puissent produire les rapports qu'ils désirent;
- b) faciliter la manipulation des demandes par des individus non-programmeurs;
- c) répondre aux différentes demandes des utilisateurs au moment approprié;
- d) exiger le moins souvent possible l'aide de programmeurs.

La réalisation de ces objectifs permettra une interaction utilisateur-ordinateur qui amènera une exploitation efficace de la banque de données et qui répondra à un grand nombre de besoins en information des utilisateurs.

6.3 Les fonctions du générateur de rapports

Les différentes fonctions à être accomplies par le générateur de rapports sont:

- a) permettre la sélection sur tous les niveaux de la banque de données;
- b) offrir différents formats de sorties des données (sous forme de liste, de tableau, de sommaire, etc.);

- c) conserver les demandes des utilisateurs pour les besoins futurs;
- d) demander le minimum d'efforts dans la formulation des demandes qui doivent être indépendantes des structures de données et des instructions à l'ordinateur.

Le générateur de rapports sera aussi un outil précieux pour l'administrateur des données dans l'accomplissement de ses fonctions. Il lui permettra de faire différentes analyses sur la banque de données pour en contrôler le comportement dans le temps.

7. LE ROLE DE L'ADMINISTRATEUR DES DONNEES

7.1 Les responsabilités de l'administrateur des données

Avec l'escalade des coûts de développement, d'implantation et d'entretien des systèmes mécanisés, l'établissement de moyens améliorés de gestion et de contrôle des données s'est avéré essentiel pour avoir un traitement efficace des données. La solution a résulté dans le développement de systèmes de gestion d'une banque de données où on a reconnu que la définition, l'organisation, la protection et la documentation de la banque de données devaient être sous la responsabilité d'un administrateur des données.

Plus spécifiquement, la nature de ces responsabilités de l'administrateur des données peut être décrite comme suit: (16)

a) La définition et l'organisation des données

L'administrateur des données doit être capable de comprendre les exigences du traitement des données de tous les utilisateurs. Pour être efficace dans ses relations interdépartementales, l'administrateur des données doit avoir la confiance de tous les utilisateurs de l'organisation et des informaticiens avec lesquels il devra communiquer avec objectivité et impartialité dans les décisions affectant plus d'un secteur de l'organisation. De plus, l'administrateur des données doit comprendre le concept de la banque de données en accord avec les objectifs à court, moyen et long terme de l'organisation pour mieux servir les demandes des utilisateurs.

L'administrateur des données a la responsabilité de définir les données actuelles ainsi que les nouvelles lorsqu'elles sont ajoutées à la banque de données, d'établir les interrelations entre les données et de documenter la banque de données.

Ceci est nécessaire pour que l'administrateur des données puisse faire connaître aux utilisateurs les données existantes qui peuvent être accédées ou desquelles l'information peut être obtenue. Si des données spécifiques n'existent pas dans la banque de données, l'administrateur des données est responsable d'identifier la source de la donnée ainsi que les exigences pour inclure la nouvelle donnée dans la banque de données.

L'administrateur des données est aussi responsable de l'organisation physique de la banque de données pour satisfaire les besoins de conservation des données et pour optimiser l'usage de la banque de données avec l'aide de programmes utilitaires et de techniques de simulation.

b) La protection des données

Une responsabilité essentielle de l'administrateur des données est de s'assurer de la protection de la banque de données contre les erreurs, la destruction et la divulgation de données confidentielles. Ceci est réalisé par le contrôle de l'accès et de la manipulation de la banque de données avec l'aide de programmes (mots de passe, identification de l'utilisateur) ou par des accès physiques restrictifs aux terminaux.

L'administrateur des données doit aussi établir et maintenir l'intégrité de la banque de données pour s'assurer que les données sont précises, complètes et à jour. Pour réaliser cela, on devra définir des systèmes communs de cueillette des données, des routines de mise à jour, des suivis communs des données erronées et la vérification sous le contrôle direct de l'administrateur des données.

Si la banque de données devait être détruite ou endommagée, il faudrait qu'elle puisse être remise en bon état. Des standards et des procédures de conservation, de récupération et de reprise sont alors nécessaires et doivent être considérés lors de la planification et de la conception du système de gestion d'une banque de données.

c) La documentation des données

La documentation est essentielle pour une administration efficace de la banque de données. L'administrateur des données est donc responsable pour développer et pour maintenir la documentation sur la description de la banque de données, les standards et les procédures, les mots de passe et les identifications des utilisateurs, les mesures de performance et d'utilisation, les procédures de conservation, de récupération et de reprise, les facilités de test et les techniques d'entraînement et d'orientation.

7.2 Les outils de l'administrateur des données

Le groupe GUIDE/SHARE a identifié différents outils pour appuyer les responsabilités de l'administrateur des données: (20)

a) le dictionnaire de la banque de données

C'est la collection de tous les éléments d'information, et de leurs attributs, de la banque de données.

b) le langage de description de la banque de données

C'est le moyen de décrire ou de définir la banque de données en termes d'interrelation entre les données, de longueur des dossiers, de facteur de regroupement, de méthodes d'accès, de volume de données, de croissance anticipée et ainsi de suite.

c) les programmes utilitaires du système

Ce sont les programmes nécessaires pour lister ou récupérer des ensembles de données ou la banque de données elle-même, pour obtenir des rapports relativement à l'intégrité et à la sécurité de la banque de données et pour fournir de l'information statistique nécessaire à l'identification des besoins de réorganisation.

d) les programmes de l'enregistrement des données

Ce sont les programmes qui fournissent de l'information sur l'allocation d'espace de la banque de données sur les unités physiques d'enregistrement.

e) les programmes utilitaires de conservation, de recouvrement et de reprise

Ce sont les programmes qui permettent la prise de copie de sécurité régulière de la banque de données ainsi que la restauration de la banque de données lors de sa destruction ou de son endommagement.

f) les facilités de vérification du système

C'est un programme qui offre les moyens de détecter les données manquantes, le report en retard de transactions, les erreurs des programmes d'application et la correction d'erreurs en retard.

g) les facilités de gestion de la banque de données

C'est un programme qui produit des statistiques sur la fréquence d'accès à la banque de données, le temps d'accès et la croissance observée.

h) les facilités de simulation de la banque de données

C'est un programme qui permet de simuler des programmes d'application pour déterminer leurs effets sur la performance de l'équipement et des programmes.

On remarque que ces outils utilisent l'ordinateur au maximum, ce qui facilite le travail de l'administrateur des données.

7.3 Les difficultés d'implantation d'un système de gestion d'une banque de données

Lors de l'implantation d'un système de gestion d'une banque de données, certaines difficultés apparaîtront et plus particulièrement:
(16)

a) la propriété des données

Dans les applications traditionnelles du traitement des données, les systèmes furent conçus et implantés pour fournir aux utilisateurs des rapports contenant des données provenant de fichiers qui furent créés, mis à jour et maintenus par cet utilisateur. Les utilisateurs, par conséquent, ont développé des attitudes possessives en regard des données contenues dans leurs fichiers et veulent les conserver jalousement.

Les données, cependant, appartiennent à l'organisation comme toute autre ressource. Plus particulièrement dans les applications utilisant une banque de données intégrée, la propriété unique d'une donnée par une personne, un groupe ou un département est impossible.

b) l'exactitude des données

Pour opérer efficacement une banque de données, les utilisateurs doivent être confiants de l'intégrité des données. La responsabilité de l'exactitude des données a toujours été assignée au département qui fournit la donnée au système. Ceci peut être un problème dans une banque de données intégrée où les utilisateurs ne connaissent pas toujours le fournisseur de la donnée et peuvent être incertains de leur exactitude.

c) la redondance des données

La redondance des données est une caractéristique qui doit être éliminer ou minimiser. Une des causes de ce problème est le manque de planification en regard des niveaux tolérables et possibles de redondance. Le problème existe dans plusieurs organisations parce qu'aucune décision n'a été prise en regard du degré de redondance acceptable.

d) la sécurité et l'interprétation des données

Avec une banque de données intégrée, un utilisateur peut avoir accès à tous les éléments d'information et par conséquent avoir accès à des données confidentielles. Ceci implique aussi que les utilisateurs peuvent interpréter différemment la même donnée.

7.4 La compétence de l'administrateur des données

L'administrateur des données doit être capable d'échanger avec différents groupes d'individus incluant les utilisateurs, les gestionnaires, les analystes, les programmeurs, les opérateurs et les fournisseurs.

Pour satisfaire ces relations interdépartementales et les responsabilités inhérentes, l'administrateur des données doit être capable de coordonner le personnel technique et non-technique. L'administrateur des données doit être orienté vers les utilisateurs.

Le personnel de support de l'administrateur des données variera selon l'importance de l'organisation, les responsabilités définies, l'implication actuelle dans le concept de la banque de données et dans les projections des applications futures de la banque de données.

8. LES "PACKAGES" DISPONIBLES

8.1 L'utilisation d'un "package" de gestion de données

Le développement d'un système de gestion d'une banque de données est une entreprise de longue haleine et peut être très dispendieuse pour l'organisation. Une telle aventure nécessite un personnel des plus compétents et des ressources monétaires importantes.

Pour réduire les risques, il existe de nombreux "packages" de gestion d'une banque de données qui peuvent être implantés rapidement et à coût raisonnable. Les principaux "packages" actuellement disponibles sont IMS-VS, TOTAL, ADABAS, SYSTEM 2000.

8.2 IMS-VS (14)

La compagnie IBM a développé le "package" IMS-VS ("Information Management System") qui rencontre les objectifs fondamentaux d'un système de gestion d'une banque de données:

a) l'indépendance des données

Avec le DL/1 ("Data Language/1"), les définitions de la banque de données, de l'ensemble des données, du dossier et du segment de données sont spécifiées et compilées à l'extérieur des programmes d'application. Durant le traitement, une application particulière est sensible seulement à son sous-ensemble de données. La modification et l'addition de données non-sensibles n'affectent pas le programme d'application.

b) la non-redondance des données

La non-redondance des données est réalisée au moyen de pointeurs qui permettent de multiples regroupements avec le même segment de données. Les structures logiques de données peuvent être organisées en niveau simple, hiérarchiquement, en réseau et en association.

c) les relations de données

Les relations de données peuvent être réalisées selon quatre méthodes d'organisation physique de la mémoire:

- le séquentiel hiérarchique
- le séquentiel hiérarchique indexé
- le direct hiérarchique
- le direct hiérarchique indexé.

La différence de base entre ces méthodes est la manière par laquelle les segments de données sont reliés et les techniques d'accès aux données.

Avec l'organisation séquentielle hiérarchique, les segments de données sont reliés par leur juxtaposition physique.

Avec l'organisation séquentielle hiérarchique indexée, le séquentiel indexé et le débordement sont utilisés. Tous les segments reliés à un maître commun sont placés dans un dossier séquentiellement indexé. Si de l'espace additionnel est nécessaire pour conserver plus de segments, un ou plusieurs blocs séquentiels de débordement sont utilisés.

Avec l'organisation directe hiérarchique, tous les segments plutôt que les blocs physiques sont reliés par des adresses directes. Cette technique permet un accès plus rapide aux segments à l'intérieur de la banque de données. L'adresse du segment maître est calculée en randomisant sur l'identificateur primaire. Les segments maîtres contiennent les pointeurs à tous les segments du prochain niveau.

Avec l'organisation directe hiérarchique indexée, l'accès au segment maître se fait séquentiellement indexé. Chacun des dossiers logiques du fichier séquentiel indexé contient l'identificateur du segment maître et un accès direct au segment maître. Aucun segment n'existe à l'intérieur du fichier séquentiel indexé. Parce que l'ensemble des données de l'index est séparé de l'ensemble des données des segments, la réorganisation de l'index peut être faite sans réorganiser les données.

d) l'intégrité et la confidentialité des données

Des programmes utilitaires sont disponibles qui permettent à l'utilisateur de spécifier le niveau de sécurité. La sécurité des terminaux et des mots de passe est supportée.

e) la performance

Trois méthodes d'accès sont disponibles pour les segments maîtres dépendant de la méthode d'organisation utilisée:

- l'accès séquentiel
- l'accès séquentiel indexé
- l'accès direct.

L'accès aux segments subordonnés (encore dépendants de la méthode d'organisation) peut être séquentiel ou direct.

f) la compatibilité

Les descriptions de la banque de données et les opérations d'entrée/sortie ne sont pas compilées dans les programmes d'application. La structure logique des données peut être réorganisée dans une organisation physique différente sans modification aux programmes. Ceci rend indépendantes les méthodes d'accès des organisations physiques de la mémoire.

Finalement IMS/VS permet l'organisation d'une banque de données pour créer, interrelier et maintenir les données d'une organisation afin qu'elle soit utilisée par une variété d'applications.

8.3 TOTAL (14)

TOTAL est un système de gestion d'une banque de données développé par CINCOM Systems Inc. de Cincinnati aux Etats-Unis. C'est un système qui accomplit la création et l'entretien d'une banque de données ainsi que la mise à jour et l'accès des données. Il fonctionne essentiellement comme une sous-routine et est incorporé au programme d'application.

TOTAL rencontre la majorité des objectifs d'un système de gestion d'une banque de données.

a) l'indépendance des données

Les définitions des données sont conservées séparément des programmes d'application. La correspondance entre les structures logiques et physiques est établie au moment de l'exécution. Des segments peuvent être ajoutés au "mapping" physique sans affecter la structure logique des données et il n'y a aucune recompilation des programmes non-affectés.

b) la non-redondance des données

La non-redondance des données est réalisée au moyen de pointeurs qui permettent d'intégrer le même segment de données à plusieurs ensembles de données. Les quatre structures logiques de IMS/VS sont supportées par TOTAL.

c) les relations de données

Les dossiers sous TOTAL sont soit des dossiers maîtres ou des dossiers de détails. Les dossiers maîtres sont organisés aléatoirement. Les dossiers de détails sont désordonnés et enchainés par des pointeurs à double sens. Le premier et le dernier dossier de détails

sont reliés au dossier maître. Les dossiers de détails peuvent appartenir à plus d'une chaîne, ce qui permet des structures de données en réseau.

d) la performance

Les dossiers maîtres sont accessibles par les méthodes d'accès direct de base de IBM. Les dossiers de détails sont accessibles séquentiellement dès que le dossier maître est en mémoire.

Lorsque des dossiers de détails sont annulés, les pointeurs des deux sens sont modifiés pour éliminer le dossier annulé et l'espace utilisé précédemment devient disponible.

Le résultat est que pour plusieurs applications, la réorganisation est éliminée.

e) la compatibilité

L'indépendance des unités physiques est réalisée avec TOTAL en utilisant d'adresse relative plutôt que l'adresse physique pour tous les dossiers de détails.

8.4 ADABAS (15)

Le système ADABAS ("Adaptable Data Base Management System") a été développé par Software AG of Darmstadt en Allemagne de l'Ouest.

ADABAS réalise les principaux objectifs d'un système de gestion d'une banque de données avec l'aide de deux principaux sous-systèmes. Un de ceux-ci est appelé le nucléus d'ADABAS et remplit les fonctions suivantes:

- a) l'enregistrement des nouvelles données;
- b) la sélection dans la banque de données à partir de certains critères;
- c) l'annulation des données;
- d) la mise à jour des données;
- e) la prévention de la perte de données à cause d'un mauvais fonctionnement des unités physiques;
- f) la gestion du flot des données entre la mémoire centrale et les unités de mémoire externes;
- g) la maintenance de statistiques sur l'usage de la banque de données.

Le deuxième sous-système d'ADABAS est composé de modules utilitaires tels que:

- a) le chargement d'un nouveau fichier;
- b) le chargement rapide pour ajouter un volume de données important à un fichier existant;
- c) l'expansion de l'espace disponible sans une réorganisation des données;
- d) les procédures de récupération de la banque de données;
- e) la présentation des statistiques d'utilisation de la banque de données;
- f) l'ajustement des paramètres contrôlant les structures de données logiques et physiques pour inclure les caractéristiques d'un fichier de données particulier;
- g) la réorganisation physique des dossiers pour augmenter la vitesse de certains traitements séquentiels spéciaux.

ADABAS est un système de gestion d'une banque de données inversé (les systèmes inversés utilisent des adresses dans des index pour localiser et accéder les données basées sur des critères de sélection).

ADABAS maintient un index pour chaque élément d'information déclaré comme un identificateur, montrant la localisation de chaque dossier contenant un identificateur d'une valeur donnée. Ces index, pour tous les identificateurs déclarés dans la banque de données, font partie de l'"associator" qui contient toutes les définitions et les relations logiques des dossiers de données. Un autre élément de l'associator est un convertisseur d'adresse qui spécifie la relation entre un numéro de séquence interne unique assigné à chaque dossier et la localisation physique sur disque. C'est par la gestion intelligente de l'associator qu'un utilisateur d'ADABAS obtient les caractéristiques de performance excellente de ce "package" pour les banques de données volumineuses.

Le fichier de données d'ADABAS est enregistré sur disque en utilisant la méthode d'accès direct de base de IBM. Les dossiers de données sont enregistrés comme des champs de longueur variable comprimés à l'intérieur des blocs, chaque dossier ayant un numéro de séquence interne unique qui lui est assigné.

En traitant une demande d'information particulière, ADABAS lie les critères de recherche spécifiés contre les champs inversés appropriés, et la stratégie de recherche la plus économique est choisie.

Parce que ADABAS emploie une banque de données inversée, les demandes complexes sont généralement traitées rapidement puisque quelques recherches seulement sont nécessaires pour charger les blocs requis de l'associator. Toutes les recherches sont habituellement faites en mémoire centrale.

8.5 SYSTEM 2000 (15)

SYSTEM 2000 est un système de gestion de banques de données généralisé développé par MRI Systems Corporation d'Austin aux Etats-Unis.

La base du SYSTEM 2000 fournit les fondements pour le développement de systèmes d'information selon les besoins des applications et des utilisateurs. SYSTEM 2000 offre de plus les options suivantes:

- un générateur de rapports
- un langage orienté pour les utilisateurs en vue d'accéder en temps réel à la banque de données
- un interface procédural pour les programmeurs, le traitement séquentiel de fichiers et l'optimisation.

La base du SYSTEM 2000 fournit à l'utilisateur un ensemble compréhensif de possibilités de gestion d'une banque de données. Elles incluent la facilité de définir de nouvelles banques de données, de modifier la définition des banques de données existantes et d'accéder et de mettre à jour les valeurs de ces banques de données.

L'utilisateur doit d'abord définir sa banque de données avec des formats de dossiers similaires à ceux utilisés dans les systèmes de gestion de fichiers traditionnels. La possibilité de modifier ce format de base lorsque la banque de données est créée est assez limitée. L'utilisateur peut alors créer sa banque de données en chargeant des champs de données qui sont dans le format de la définition des données établie précédemment.

Les composantes de base de la définition de la banque de données sont les éléments d'information et les groupes à répétition. Les valeurs sont conservées dans les éléments d'information. Les groupes à répétition décrivent une structure pour conserver des ensembles multiples de valeurs de données et aussi servir pour relier les niveaux hiérarchiques de la définition.

SYSTEM 2000 permet à un fichier d'être structuré d'une manière telle qu'un ou plusieurs champs dans la banque de données peuvent être indexés et que les éléments d'information associés dans les groupes à répétition sont accédés sans passer tous les groupes à répétition reliés. Ces index sont conservés séparément des données. Donc, tous les processus de sélection travaillent avec les pointeurs des index plutôt qu'avec les données. Lorsque le processus de sélection est complété, les pointeurs des index qui restent pointent directement aux données à être accédées. Le chaînage de dossiers des données ne se fait jamais. De plus, toutes les données pour une entrée logique sont normalement conservées ensemble et peuvent être accédées avec une recherche.

SYSTEM 2000 utilise des fichiers à réseau, ces derniers étant partiellement inversés pour un accès en temps réel efficient; il laisse la détermination du degré d'inversion à l'utilisateur.

Les valeurs de chaque élément d'information et des dossiers logiques peuvent varier en longueur. L'utilisateur peut spécifier sans restriction quels éléments de la banque de données doivent être inversés et devenir des champs d'identificateur, et quelle relation hiérarchique un élément d'information aura avec les autres éléments d'information de la banque de données.

8.6 Le choix d'un "package"

L'acquisition de l'un ou l'autre de ces "packages" doit être précédée d'une étude plus approfondie de chacun de ceux-ci pour identifier lequel répondra le mieux aux besoins de l'organisation. Aucun de ces "packages" ne répondant intégralement aux objectifs d'un système de gestion d'une banque de données, il faudra cerner les avantages et les inconvénients de chacun de ceux-ci en relation avec les besoins de l'organisation.

Quant au coût d'acquisition de l'un ou l'autre de ces "packages", celui-ci est relativement élevé (entre \$50,000. et \$200,000.) mais il peut amener de réelles économies lors de leur implantation.

9. POUR ARRIVER À L'IMPLANTATION D'UNE BANQUE DE DONNEES

La décision d'implanter un système de gestion de banque de données ne peut être prise sans que les étapes suivantes soient franchies de façon satisfaisante:

a) Le regroupement des utilisateurs et du personnel de système

L'intégration d'un système de gestion d'une banque de données au système d'information de l'organisation ne peut être possible sans une communication adéquate entre les utilisateurs du système d'information et le personnel affecté à son développement et à son entretien. Il est donc important de former un groupe de travail responsable de l'implantation d'une banque de données formé de représentants des utilisateurs et de l'informatique sous autorité directe du management.

b) La revue et la critique des systèmes existants

On devra d'abord identifier les variables-clés de l'organisation sur lesquelles se prennent généralement les décisions. Les variables-clés permettront d'identifier les éléments d'information de base de la banque de données.

Le processus d'identification des éléments d'information se poursuivra en examinant les fichiers de transactions et les rapports existants.

En examinant l'usage des éléments d'information des systèmes existants, on identifiera des inconsistances et de la redondance dans les données. La variation dans la longueur des champs, dans les règles d'édition et dans l'interprétation des données contribue à l'inconsistance des données dans les fichiers. Les données conservées dans plusieurs fichiers exigent une mise à jour redondante.

La reconnaissance des conflits entre les éléments d'information fournit une bonne base pour la planification de la banque de données.

Cette analyse permettra de plus d'identifier les responsabilités des données ou de relever les conflits pouvant exister à ce niveau.

Finalement, des éléments d'information additionnels et des besoins de contrôle futur seront aussi identifiés lors de cette revue des systèmes existants.

c) La conception générale de la banque de données

Les objectifs du système sont essentiels pour la conception de la banque de données. Aussi ceux-ci doivent-ils être identifiés par le management et les utilisateurs en collaboration avec le personnel du système.

En se basant sur ces objectifs et sur la revue et la critique des systèmes existants, on définira la structure logique des données qui devra être supportée par la banque de données en identifiant les interrelations possibles entre les divers ensembles d'éléments d'information.

Les frontières d'un système d'information s'étendant au-delà du traitement sur ordinateur, on devra identifier le flot d'information complet à partir de la cueillette des données jusqu'au résultat final. Ces frontières incluent de plus les actions et les responsabilités de tous les secteurs de l'organisation affectée par le système.

La conception générale du système devra donc inclure le flot général des données, les responsabilités, les exigences du traitement et les contrôles.

d) La confrontation avec les objectifs

Les résultats de la conception générale de la banque de données devront être confrontés avec les objectifs établis précédemment. Si ceux-ci ne peuvent être réalisés, la conception générale de la banque de données devra être revue.

e) Le choix du support de la banque de données

Une analyse des "packages" disponibles doit être entreprise lesquels seront confrontés avec les exigences de la conception générale du système. Pour chaque "package" analysé, on devra identifier le pour et le contre du choix dudit "package".

Avant de rejeter le choix d'un "package" pour supporter la banque de données de l'organisation, le management doit évaluer sérieusement les implications du développement interne d'un système de gestion de banque de données. Avant de prendre une telle décision, on devra considérer les points suivants:

- la compétence du personnel de système;
- la durée du développement du nouveau système;
- le coût du développement du nouveau système;
- la finalité des standards et des mécanismes actuels de la documentation;
- le contrôle du management sur le développement du nouveau système;

- le risque et la probabilité de réussite.

Ces étapes sont importantes pour la réussite de l'implantation d'un système de gestion d'une banque de données. Aussi le management doit-il en contrôler l'évolution et exiger des rapports progressifs pour connaître continuellement l'état de la situation.

GLOSSAIRE

1. Administrateur des données ("Data Base Administrator")

Gestionnaire d'une équipe responsable de l'administration d'une banque de données qui est une activité couvrant la définition, l'organisation, la protection et la documentation des données de la banque de données.

2. Aire de mémoire ("Space extent")

Portion physique continue d'une mémoire externe de l'ordinateur.

3. Attribut ("Attribute")

Caractéristique d'un élément d'information telle que sa longueur, son format, ses limites, etc..

4. Banque de données ("Data Bank")

Collection d'un ou plusieurs ensembles de données reliés logiquement.

5. Banque de données intégrée ("Integrated Data Bank")

Banque de données qui contient les données pertinentes à plus d'une fonction de l'organisation.

6. Débordement ("Overflow")

Aire de mémoire qui recueille les dossiers ne pouvant être enregistrés dans l'aire de mémoire prévue faute d'espace disponible.

7. Dictionnaire des données ("Data Base Dictionary")

Ensemble des définitions des éléments d'information d'une banque de données et de leurs attributs.

8. Donnée ("Data")

Fait brut représenté par un symbole.

9. Dossier logique des données ("Logical Data Base Record")

Regroupement d'un ou plusieurs segments de données reliés logiquement soit de façon contiguë à l'intérieur d'un bloc physique ou soit par des pointeurs.

10. Element d'information ("Data Item", "Data Element")
Unité la plus petite d'une donnée définie dans la banque de données.
11. Ensemble de données ("Data set")
Ensemble relié de dossiers logiques de la banque de données.
12. Entité ("Entity")
Personne, un évènement, un objet, un endroit d'intérêt pour l'organisation.
13. Fichier ("File")
Ensemble de données contenant des dossiers dans lesquels les segments de données sont contigus à l'intérieur des blocs physiques reliés.
14. Générateur de rapports ("Report generator")
Un ou plusieurs programmes établis pour produire rapidement et facilement les différents rapports désirés par les utilisateurs.
15. Identificateur primaire ("Primary key")
Champs de contrôle ou l'identificateur d'un dossier identifiant uniquement un dossier logique.
16. Identificateur secondaire ("Secondary key")
Element d'information descriptif appartenant à un segment de données et sur lequel on désire avoir accès.
17. Indépendance des données ("Data Independence")
Séparation des définitions du dossier logique et du dossier physique de telle sorte que les programmes d'applications ne sont pas dépendants des unités physiques d'enregistrement et des méthodes d'accès.
18. Index ("Index")
Table reliée à une procédure qui accepte comme entrée les valeurs de certains attributs et qui donne comme sortie la donnée permettant de localiser rapidement le dossier ou les dossiers possédant la valeur de l'attribut.

19. Information ("Information")

Ensemble de données qui ont été traitées dans une forme qui est significative au récepteur et qui est d'une valeur réelle ou perçue dans les décisions courantes ou à venir.

20. Intégrité des données ("Data Integrity")

Qualité des symboles représentant, de façon conforme, ces faits bruts. C'est le fait que les données doivent être protégées contre une invalidation accidentelle ou délibérée de sa valeur ou de ses relations avec d'autres données.

21. Liste filée ("Threaded List")

Technique de structuration de liste dans laquelle les pointeurs sont impliqués dans les dossiers logiques de données d'une partition, reliant ensemble les dossiers dans une chaîne. Pour chaque partition, il y a une entrée dans un répertoire. L'entrée est un pointeur au premier dossier dans la partition. Dans chaque dossier il y a un pointeur au prochain dossier dans la chaîne.

22. Liste inversée ("Inverted List")

Technique de structuration de liste dans laquelle tous les pointeurs sont extraits des dossiers de la partition et enregistrés dans un index. Un répertoire est maintenu avec un pointeur pour chaque partition, pointant à l'index approprié.

23. Méthode d'accès ("Access Method")

Moyen par lequel on accède aux segments de données dans la structure d'enregistrement physique.

24. Organisation d'un fichier ("File organisation")

Correspondance existant entre la structure des données et la structure d'enregistrement physique.

25. Partition de fichier ("File Partition")

Série de dossiers reliés logiquement associés par des pointeurs, plutôt qu'être enregistrés de façon contiguë, de telle manière qu'un sous-fichier relativement petit est accessible.

26. Pointeur ("Pointer", "link address")
Adresse utilisée pour permettre la communication entre les dossiers logiques d'une partition.
27. Programme d'application ("Application program")
Programme écrit pour le traitement de données d'un utilisateur.
28. Randomiser ("Randomize")
Convertir un nombre selon des règles définies à l'avance pour obtenir l'adresse d'un dossier.
29. Redondance ("Redondancy")
Situation dans laquelle une donnée se retrouve en plusieurs exemplaires dans plusieurs fichiers différents.
30. Schéma ("Schema")
Description complète de tous les éléments d'information d'une banque de données.
31. Segment de données ("Data Segment")
Champs de données de longueur fixe contenant un ou plusieurs éléments d'information reliés logiquement et physiquement.
32. Sous-schéma ("Subochema")
Sous-ensemble logique du schéma qui est accédé par un ou plusieurs programmes spécifiques.
33. Structure de données de niveau simple ("Single Level Data Structure")
Structure logique de données ne contenant qu'un segment maître.
34. Structure de données en association ("Associative Data Structure")
Structure logique de données dans laquelle les segments de données sont reliés de telle sorte que chacun contient une donnée qui est identique.

35. Structure de données en réseau ("Network Data Structure")

Structure logique de données dans laquelle un segment de détail est partagé par plus d'un segment maître.

36. Structure de données hiérarchique ("Hierarchial Data Structure")

Structure logique de données dans laquelle il y a un segment de données portant une relation hiérarchique supérieure à tous les segments du prochain niveau. Dans cette relation, le segment supérieur est appelé "segment maître" et le subordonné est appelé "segment de détail". Le segment unique du niveau le plus élevé est appelé le "segment racine".

37. Structuration de liste ("List structuring")

Technique d'organisation d'un fichier pour un accès aléatoire à des partitions logiques.

38. Structure logique de données ("Logical Data Structure")

Ensemble de segments de données reliés qui servent de modèle au dossier logique d'une banque de données.

39. Système ("System")

Ensemble d'éléments en interaction qui opèrent ensemble pour atteindre un objectif.

40. Système d'information pour la gestion ("Management Information System")

Méthode organisée pour fournir de l'information passée, présente et projetée reliées aux opérations internes et à l'environnement. Il supporte les fonctions de planification, de contrôle et d'opération d'une organisation en fournissant une information uniforme au moment approprié pour assister le décideur.

41. Système de gestion d'une banque de données ("Data Base Management")

Système de traitement de données accomplissant les fonctions de manipulation de données pour une banque de données.

42. Système de gestion de fichier ("File Management System")

Système de traitement de données accomplissant les fonctions d'entrée, de validation, de création de fichier, de mise à jour de fichier, d'extraction de données et de présentation de rapports.

43. Utilisateur ("End-user", "user")

Personne située à l'un ou l'autre des niveaux de l'organisation dont le travail nécessite l'utilisation d'une banque de données de façon plus ou moins prédominante.

BIBLIOGRAPHIE

1. Anderson Clayton Foods, Preliminary Data Dictionary/Directory System Planning Guide, Manual DD/D-1-X.1, 1973.
2. Bruun, Roy, Infosystems, "Ingredients of a Data Base", Décembre 1973.
3. Burch, John G. et Felix R. Strater, Information Systems: Theory and Practice, Hamilton Publishing Company, Santa Barbara, 1974.
4. Coleman, Raymond J. et M.J. Riley, MIS: Management Dimensions, Holden-Day Inc., San Francisco, 1973, pp. 3-32.
5. Curtice, Robert M., Datamation, "Data Independence in Data Base Systems", Avril 1975.
6. Curtice, Robert M., Datamation, "Some Tools for Data Base Development", Juillet 1974.
7. Davis, Gordon B., Management Information Systems: Conceptual Foundations, Structure and Development, McGraw-Hill Book Company, New-York, 1974.
8. Dearden, John, Harward Business Review, "MIS is a Mirage", Janvier-Février 1972.
9. Evans, Ronald W., EDP In-Depth Reports, "Data Base Organisation for On-Line Systems", Vol. 1 No 1, Septembre 1971.
10. Evans, Ronald W., EDP In-Depth Reports, "Data Base Management Systems", Vol. 1 No 2, Octobre 1971.
11. Evans, Ronald W., EDP In-Depth Reports, "File Management Systems", Vol. 1 No 3, Novembre 1971.
12. Evans, Ronald W., EDP In-Depth Reports, "Fast Response Information Retrieval Systems", Vol. 1 No 4, Décembre 1971.
13. Evans, Ronald W., EDP In-Depth Reports, "Current Developments in Data Base Management", Vol. 2 No 5, Janvier 1973.
14. Evans, Ronald W., EDP In-Depth Reports, "New Software for Data Base Management: Part 1", Vol. 3 No 1, Septembre 1973.
15. Evans, Ronald W., EDP In-Depth Reports, "New Software for Data Base Management: Part 2", Vol. 3 No 2, Octobre 1973.

16. Evans, Ronald W., EDP In-Depth Reports, "The Data Base Administrator: Part 1", Vol. 3 No 12, Août 1974.
17. Evans, Ronald W., EDP In-Depth Reports, "The Data Base Administrator: Part 2", Vol. 4 No 1, Septembre 1974.
18. Ficarra, Anthony, Ideas for Management, "Data Base Systems Design".
19. Galley, Thomas A., Journal of Systems Management, "An approach to Data Base Design", Février 1969.
20. GUIDE DATA DICTIONARY/DIRECTORY PROJECT, Requirements for the Data Dictionary/Directory within the GUIDE/SHARE Data Base Management System Concept, Novembre 1974.
21. GUIDE Basic Data Base Requirements Project, Basic Requirements for a Data Base Management System, Février 1973.
22. Head, Robert V., Journal of Systems Management, "Structuring the Data Base for Management Information Systems", Janvier 1969.
23. Heydon, Michael J., Canadian Data systems, "How to Prepare for Effective Data Management".
24. Horton, Forest W., How to Harness Information Resources: A Systems Approach, Association for Systems Management.
25. IBM Corporation, Information Management System Virtual Storage (IMS/VS) - General Information Manual GH20-1260-0, Janvier 1973.
26. Keysor, Frederic, Data Management, "The Managed Data Base", Mai 1974.
27. Pohara, Charanzit S., Journal of Systems Management, "A New Approach to Management Information Systems - Part 1", Juillet 1971.
28. Pohara, Charanzit S., Journal of Systems Management, "A New approach to Management Information Systems - Part 11", Août 1971.
29. MacLachean, Ian, Canadian Data systems, "MIS: The Impossible Dream?", Juin 1973.
30. Martin, James, Computer Data-Base Organization, Prentice-Hall, Englewood Cliffs, 1975.

31. Nolan, Richard L., Harvard Business Review, "Computer Data Bases: The Future Is Now", Septembre-Octobre 1973.
32. Patterson, Albert C., Datamation, "Data Base Hazards", Juillet 1972.
33. Price, Gerald F., Data Management, "The Ten Commandments of Data Base", Mai 1972.
34. Romberg, Bernhard W., Infosystems, "Data Bases: There Really Is a Better Way to Manage Your Files", Mai 1973.
35. Schubert, Richard F., Datamation, "Directions in Data Base Management Technology", Septembre 1974.
36. Schubert, Richard F., Datamation, "Basic Concepts in Data Base", Juillet 1972.
37. Traver, Cheryl M., Data Bases: Uncontrollable or Uncontrolled?, Stanford Center for Information Processing, Stanford.
38. Traver, Cheryl M., Data Base Administrator - - The Emerging Position, College and University Machine Records Conference, Mai 1973.
39. Will, Hart H., Journal of Systems Management, "MIS: Mirage or Mirror Image?", Septembre 1973.

MINISTERE DES TRANSPORTS



QTR A 101 791