



# les fichiers

pratique et choix de l'organisation  
des données informatiques

IST 215

par

**Claude JOUFFROY**

Ingénieur, responsable d'organisation

et

**Charles LÉTANG**

Maître-assistant au Département d'Informatique de l'I.U.T. d'Orsay I

2<sup>e</sup> édition

BIBLIOTHEQUE DU CERIST

IST 215

phase spécialité

**DUNOD**

**informatique**

© BORDAS, Paris, 1977 — 021 677 0205

I.S.B.N. 2-04-003123-5

“ Toute représentation ou reproduction, intégrale ou partielle, faite sans le consentement de l'auteur, ou de ses ayants-droit, ou ayants-cause, est illicite (loi du 11 mars 1957, alinéa 1<sup>er</sup> de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal. La loi du 11 mars 1957 n'autorise, aux termes des alinéas 2 et 3 de l'article 41, que les copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective d'une part, et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration ”

# préface

*Si l'Informatique est définie comme la science du traitement rationnel de l'information, considérée comme le support des connaissances, il est paradoxal de constater que la majorité des ouvrages qui lui ont été consacrés jusqu'à présent traitent relativement peu de la nature de cette information, et encore moins de la façon dont elle doit être organisée pour pouvoir être manipulée par un automate déterminé selon une procédure fixée.*

*Certes, les organisations de données dont l'analyste peut disposer sont étroitement liées au langage de programmation qui sera utilisé et encore plus au matériel sur lequel sera réalisée l'exploitation.*

*Aussi est-il couramment admis que l'exposé des différentes méthodes possibles de rangement des informations fait partie de la documentation technique des constructeurs d'ordinateurs.*

*Mais si les systèmes de Gestion des Entrées-Sorties actuellement fournis avec les équipements répondent généralement à la plupart des besoins d'un concepteur d'application, il s'avère qu'il est bien souvent utile de développer des organisations non supportées par le système d'exploitation.*

*Il en est ainsi dès que l'on procède à des traitements relativement complexes, que l'on désire améliorer soit le temps d'exploitation, soit le taux d'encombrement des mémoires externes, et davantage encore lorsque l'on veut équilibrer la répartition des charges entre l'unité centrale et les unités d'échange.*

*Cette nécessité se retrouve également lorsque l'on utilise des petits systèmes informatiques pour lesquels les procédures de Gestion des Entrées-Sorties sont en général peu développées.*

*Aussi nous a-t-il semblé nécessaire de rassembler dans un ouvrage les notions essentielles relatives à cette question, afin que chaque informaticien*

## PRÉFACE

*puisse disposer, condensés en quelques pages, des principes de base de l'organisation des données et des critères lui permettant de procéder au choix de l'organisation la mieux adaptée à une application.*

*Les auteurs ont volontairement donné à cet ouvrage un caractère très pratique, afin que les idées qu'il renferme puissent être utilisées immédiatement. Nous espérons qu'il rendra service à beaucoup d'informaticiens, mais aussi qu'il fera comprendre aux utilisateurs de l'informatique la complexité du choix de l'organisation la mieux adaptée à un ensemble de données et la nécessité pour l'analyste d'être informé, aussi complètement que possible, à la fois de leurs besoins, de l'évolution de ces besoins et des spécifications des données.*

*C'est la condition à laquelle l'informatique pourra mettre en place des applications qui donnent réellement satisfaction aux utilisateurs et ne se périment pas prématurément.*

Jacques GIROIRE

Président-Directeur général de GMV Conseil

# Sommaire

Pages

## I - généralités sur les fichiers

1 — Notion de fichier	1
2 — Caractéristiques des fichiers	3
2.1 - Définitions	3
21.1 Le caractère	3
21.2 La zone	4
21.3 L'enregistrement	4
21.4 Le fichier	5
2.2 - Caractéristiques d'utilisation des fichiers	8
22.1 Activité d'un fichier	8
22.11 Taux de consultation	8
22.12 Fréquence de consultation	9
22.13 Fréquence par tranches d'indicatif	10
22.14 Fréquence par article	11
22.15 Taux de renouvellement — Fréquence	11
22.16 Stabilité	11
22.2 Volume d'un fichier	11
22.3 Accroissement et vitesse de croissance	12
3 — Typologie des fichiers	13
3.1 - Les fichiers permanents (ou de base)	13
3.2 - Les fichiers mouvements	14
3.3 - Les fichiers de manœuvre et les fichiers intermédiaires	15
33.1 Les fichiers de manœuvre	16
33.2 Les fichiers intermédiaires	16
3.4 - Fichiers archives, fichiers historiques	16

	Pages
3.5 - Autres fichiers	17
35.1 Les fichiers entrées	17
35.2 Les fichiers résultats	17
35.3 Les tables	17
4 — Principaux traitements des fichiers	18
4.1 - Les traitements fonctionnels	18
41.1 Prise en compte et validation des informations :	
Les contrôles	19
41.2 Élaboration des informations	20
41.21 Réunion	20
41.22 Éclatement	21
41.3 Valorisation	21
41.4 Tri et fusion	21
41.41 Tri	21
41.42 Fusion	22
41.5 Mise à jour	22
41.6 Extraction	23
4.2 - Les traitements de servitude	23
42.1 Copie	23
42.2 Mémorisation temporaire	23
42.3 Création des fichiers annexes	24

## II - rappels technologiques sur les supports

1 — Caractéristiques générales des mémoires auxiliaires	26
2 — Bande magnétique	27
2.1 - Description	27
2.2 - Principe de l'enregistrement	27
2.3 - Les modes d'enregistrement	28
23.1 Le mode NRZ	29
23.2 La modulation de phase (PE)	29
23.3 Comparaison des deux modes d'enregistrement	30
2.4 - Format des enregistrements	30
24.1 Notion de bloc	30
24.2 Format des enregistrements NRZ	31
24.3 Format des enregistrements PE	31
2.5 - Capacité de stockage	32
25.1 Densité d'enregistrement	32
25.2 Capacité théorique d'une bande	32
25.3 Capacité réelle de stockage	33
2.6 - Caractéristiques des dérouleurs de bandes	35
3 — Les mémoires adressables	37
3.1 - Le disque magnétique	38
31.1 Description	38
31.2 Format d'une piste	40
31.3 Types d'unités de disques	43

# TABLE DES MATIÈRES

	Pages
31.4 Commandes	43
31.41 Recherche	43
31.42 Lecture	44
31.43 Écriture	45
31.5 Caractéristiques des unités de disques	45
31.6 Capacité de stockage	48
3.2 - Le tambour magnétique	49
32.1 Description	49
32.2 Caractéristiques	49
3.3 - Les feuillets magnétiques	50
33.1 Description	50
33.2 Temps d'accès aux informations	54
33.3 Caractéristiques	55
3.4 - Format des enregistrements	56
34.1 Enregistrements fixes non bloqués	56
34.2 Enregistrements fixes bloqués	57
34.3 Enregistrements variables non bloqués	57
34.4 Enregistrements variables bloqués	57
34.5 Enregistrements indéfinis	58
34.6 Enregistrements étendus	58
3.5 - Comparaison des mémoires adressables	59

## III - l'organisation des fichiers

1 — Structure générale des fichiers ✓	60
1.1 - L'enregistrement (ou enregistrement logique)	60
1.2 - Types d'enregistrement	62
12.1 Enregistrements de longueur fixe	62
12.2 Enregistrements de longueur variable	64
12.3 Enregistrements de longueur indéfinie	65
12.4 Enregistrements variables étendus	65
1.3 - L'enregistrement physique ou bloc	65
13.1 Bloc fixe	65
13.2 Bloc variable	66
13.3 Détermination du blocage	66
1.4 - Le volume	67
1.5 - Le fichier	68
15.1 Fichier monovolume	68
15.2 Fichier multivolume	69
15.3 Volume multifichier	69
2 — Protections et sécurité sur les fichiers	69
2.1 - Les sécurités hardware	70
2.2 - Les sécurités software	71
2.3 - L'organisation de l'exploitation	73
2.4 - Restauration des fichiers	74
3 — Principales méthodes d'organisation	77
3.1 - Organisation séquentielle physique ✓	78

	Pages
31.1 Bande magnétique	78
31.11 Création du fichier	78
31.12 Consultation	78
31.13 Mise à jour	78
31.14 Coût de la maintenance	79
31.15 Conclusions	80
31.2 Les supports adressables	80
31.21 Création du fichier	80
31.22 Consultation	80
31.23 Mise à jour	81
31.24 Coût de la maintenance	81
31.25 Conclusions	81
3.2 - Organisation séquentielle logique	82
32.1 Création du fichier	83
32.2 Consultation	83
32.3 Mise à jour	83
32.31 Création d'un enregistrement	83
32.32 Suppression d'un enregistrement	85
32.4 Coût de la maintenance	85
32.5 Conclusions	85
32.6 Compression des données	86
32.61 Création du fichier	86
32.62 Création d'enregistrement	89
32.63 Suppression d'enregistrement	90
32.64 Modification d'enregistrement	91
32.65 Contraintes	91
3.3 - Organisation séquentielle indexée	91
33.1 Création du fichier	91
33.2 Consultation	93
33.3 Mise à jour	94
33.31 Modification	94
33.32 Suppression	94
33.33 Création d'un enregistrement	94
33.4 Coût de la maintenance	96
Planches I à VIII	
3.4 - Organisation adressée	97
34.1 Principes	97
34.2 Problèmes posés par l'organisation adressée	97
34.3 L'adressage	99
34.31 Description d'une adresse	100
34.32 Description des données	100
34.33 Types d'adressage	100
34.34 L'adressage relatif	103
34.4 Production de nombres aléatoires	106
34.41 Transformation par division	107
34.42 Extraction simple	108
34.43 Extraction du carré	108
34.44 Pliage	108
34.45 Traduction simple	108
34.46 Changement de base de numération	108
34.47 Traduction par table	109
34.48 Clés à zones multiples	109
34.49 Analyse des fréquences	109

TABLE DES MATIÈRES

	Pages
34.5 Choix d'une « bonne » transformation	111
34.51 Approche générale du problème	111
34.52 Test du nombre de sommets	112
34.53 Loi de Poisson sur les synonymes	115
34.6 Les collisions	116
34.61 Un seul enregistrement par case	117
Planches IX à XVIII	
34.62 Plusieurs enregistrements par case	121
34.7 Amélioration des performances de l'adressage calculé	122
34.71 Utilisation de mémoires hiérarchisées	122
34.72 Utilisation successive de plusieurs transformations	122
34.73 Critères d'utilisation	123
4 — Langages et organisations	124
4.1 - COBOL	124
41.1 Généralités	124
41.2 Organisation séquentielle indexée	125
41.3 Organisation adressée	126
41.31 Accès direct	126
41.32 Accès séquentiel	129 ✓
4.2 - COBOL — ANS	129
4.3 - PL/1	132
43.1 Généralités	132
43.2 Organisation séquentielle indexée	134
43.3 Organisation adressée	134
43.31 Organisation REGIONAL (1)	134
43.32 Organisation REGIONAL (2)	135
43.33 Organisation REGIONAL (3)	135
4.4 - FORTRAN	136
44.1 Généralités	136
44.2 L'organisation adressée	136
5 — Ensembles structurés de données	137
5.1 - Les files	137
5.2 - Les tableaux	138
5.3 - Les arborescences	139
5.4 - Les piles	142
5.5 - Les tables	143
6 — Les bases de données	149
6.1 - Introduction	149
6.2 - Les fichiers chaînés	150
62.1 Le chaînage	150
62.2 Organisation du fichier de base	150
62.3 Fichiers d'appel multi-critères	151
62.4 Création du fichier de base	151
62.5 Consultation du fichier	151
62.6 Création d'enregistrements	153
62.7 Mise en place d'un critère	153
62.8 Suppression d'un enregistrement	153

	Pages
6.3 - Traitements des arborescences	153
63.1 Représentation d'une arborescence	155
63.2 Mise à jour d'une arborescence	157
63.3 Applications	159
63.4 Généralisation	159
6.4 - Les fichiers inversés	160
64.1 Généralités	160
64.2 Maintenance d'un fichier inversé	161
7 — Les systèmes de gestion de fichiers	162
7.1 - Intérêt de ces systèmes	163
7.2 - Classification des systèmes	163
7.3 - Choix d'un système	164
7.4 - Mise en place	166
7.5 - Présentation des principaux systèmes	166
75.1 Systèmes généralisés d'interrogation de fichiers et de génération d'états	167
75.2 Manipulation de fichiers	169
75.3 Élaboration de tableaux	170
75.4 Gestion des bases de données	170

## IV - le choix d'une organisation

1 — Position du problème	172
2 — Les paramètres du choix	173
2.1 - Les données	173
2.2 - Les traitements	176
2.3 - Les contraintes externes	178
3 — Le choix	180
3.1 - Une méthode	180
31.1 Généralités	180
31.2 Première phase	180
31.3 Deuxième phase	181
3.2 - Application de la méthode à un problème de facturation	183
32.1 Première phase	184
32.2 Deuxième phase	186
3.3 - Autre exemple : validation des mouvements en mode conversationnel	188
33.1 Rappels sur la normalisation des procédures de contrôle	188
33.11 Principes d'analyse des contrôles	188
33.12 Critères intervenant dans un contrôle	189
33.13 Déclenchement d'un contrôle	191
33.2 Définition du problème	191
33.3 Organisation des fichiers	194
33.31 Le fichier des segments	194
33.32 Le fichier PARAMAT	195
33.33 Le fichier des procédures	196
Index	197