

Contents

Preface

Chapter 1 Definitions Related to Data

1.1	Definitions of data groups	1
1.1.1	Databases	1
1.1.2	Comparison of the database and the classical approaches	6
1.1.3	Logical file	7
1.1.4	Logical record	7
1.1.5	Physical files and records	8
1.1.6	Field	8
1.2	Definitions related to processing	8
1.2.1	Disk page	9
1.2.2	Block and buffer	9

Chapter 2 Data Representation and Storage Methods

2.1	Full-length storage	10
2.2	Representing data with codes	10
2.3	Vocabulary and alphabet	11
2.4	Base	12
2.5	Data packing; multiple bases	13
2.5.1	Packing	14
2.5.2	Interpretation of the packing formula	14
2.5.3	Unpacking	15
2.5.4	Further packing	15
2.5.5	Further unpacking	17
2.5.6	Storage economy	17
2.6	Packing using multiple words	18
2.6.1	Technique	18
2.7	When is data packing worthwhile?	20
2.7.1	The cost of packing	20
2.7.2	Processing time economies when sorting or searching	20
	2.7.2.1 Sorting time savings	21
	2.7.2.2 Retrieval time savings	21
2.7.3	What fields should be packed?	21
2.7.4	'Secret' packing — Encryption	22
2.7.5	Dynamic password	22

2.8	How to reduce the cost of data packing: use of bases of the form 2 ⁿ	23
2.8.1	Choosing the bases	23
2.8.2	Packing and unpacking	23
2.8.3	Advantages of this method	23

Chapter 3 Disk File Structures

3.1	Fundamental operations using a database	24
3.1.1	Initial creation	25
3.1.2	Record addition/subtraction	25
3.1.3	Individual field content modification	26
3.1.3.1	First MOD definition extension	27
3.1.3.2	Second MOD definition extension	28
3.1.3.3	Third MOD definition extension	28
3.1.4	Data selection (retrieval)	29
3.1.4.1	Primary definition	29
3.1.4.2	Secondary definition	29
3.1.4.3	Comments about standard selection outputs	31
3.1.5	Database clean-up and restructuring	32
3.1.6	Definition of batch processing, transaction processing, real-time processing	32
3.2	Criteria for choosing a structure	33
3.2.1	The type of database	33
3.2.1.1	The master file type	33
3.2.1.2	The history file type	34
3.2.1.3	The open-item file type	34
3.2.1.4	Complex structures	34
3.2.1.5	Comparison of operations on the three main types of database	34
3.2.2	The amount of data to be stored	35
3.2.3	The processing required	36
3.2.4	The logical structure of the data	36
3.3	Structures for master file databases	36
3.3.1	Sequential 'stack'	37
3.3.2	Sorted sequential file: index file	38
3.3.2.1	Operations	38
3.3.2.2	Remarks on this structure	39
3.3.2.3	Definitions	40
3.3.3	Hierarchical index	40
3.3.4	Index with partially filled blocks	41
3.3.5	Indexed structures	42
3.3.5.1	Existence (bit) tables	43
3.3.5.2	Operations on indexed structures	46
3.3.6	Indexed sequential structure	47
3.3.7	Computed address ('hash coding')	48
3.3.7.1	Addressing algorithms for numeric and alphanumeric keys	49
3.3.7.2	Structures suitable for hash coding	54

3.3.7.3	Fundamental operations on databases with hash coding	56
3.3.8	How to choose a master file structure	57
3.3.8.1	File Types	57
3.3.8.2	File Size	57
3.3.8.3	Operations	58
3.3.9	Examples of customer file structure	61
3.3.9.1	Is the file a master file?	62
3.3.9.2	File sizes	62
3.3.9.3	Choice between indexed structure and hash code structure	62
3.3.9.4	Detailed study of structures	63
3.4	Structures for history databases	68
3.4.1	Sample history databases	68
3.4.1.1	The 'invoice history' file	68
3.4.1.2	The 'stock history' file	68
3.4.1.3	The ledger	68
3.4.2	Fundamental operations on history databases	69
3.4.3	Database levels—hierarchies	69
3.4.4	Structures, lists and selection indexes	71
3.5	Structures for open item databases	72
3.5.1	Sample open item database	72
3.5.2	Fundamental operations and appropriate structure	73
3.6	Complex structure—open content databases	74
3.6.1	Evolution of a database	74
3.6.2	Addition of new fields to a database	75
3.6.2.1	Addition in an existing record	75
3.6.2.2	Using parallel files	76
3.6.2.3	Using supplementary files and pointers: total opening	76
3.6.3	Open content databases	78
3.6.3.1	Definitions	78
3.6.3.2	Linkage structure	84
3.6.3.3	Storage structure, FIFO and LIFO lists	87
3.6.3.4	Operations on open content databases	89
3.6.3.5	Example of a commercial database	95

Chapter 4 Database Development Techniques

4.1	General principles	108
4.1.1	Identifying the fundamental operations	108
4.1.2	The nucleus of a database	109
4.1.3	Three important features: accessibility, generality, flexibility	110
4.1.3.1	Accessibility	111
4.1.3.2	Generality	111
4.1.3.3	Flexibility	113
4.1.4	Unifying the processing procedures; sample data manipulation language	114
4.1.4.1	Selection language—language objectives	116

4.1.4.2	Syntactic and semantic peculiarities	117
4.1.4.3	Pictorial representation of selection syntax	118
4.1.4.4	Recursive definition	119
4.1.4.5	Points concerning graphical representation	119
4.1.4.6	Notes on the fundamental operations other than selection	123
4.1.5	Controlling access to a database: the 'window' principle	125
4.1.5.1	Creation of a regulating/optimizing window	126
4.1.6	Integrating several databases	127
4.1.6.1	Definition	127
4.1.6.2	Redundancy problem	128
4.1.6.3	Fundamental operations on integrated databases	128
4.1.6.4	Search path and transfers	128
4.1.6.5	Conclusion	129
4.2	Some programming techniques	130
4.2.1	Binary search	130
4.2.2	Improved binary search	130
4.2.3	How to design a data selection program	131
4.2.3.1	General architecture	132
4.2.3.2	Design of comparison and output modules: use of the 'computed GO TO'	133
4.2.3.3	Notes on selection programs	135
4.2.4	Automatic reporting and statistics	135
4.2.4.1	Statistics program architecture	136
4.2.4.2	Using standard (SELF) files for relational processing	136
4.3	Database development procedure	138
4.3.1	Specifications	138
4.3.2	Defining the data to be used: data dictionary	139
4.3.3	Defining the necessary hardware	140
4.3.4	General system design	140
4.3.4.1	Defining logical data groups	141
4.3.4.2	Defining physical files	141
4.3.4.3	Appropriate data representation and packing	141
4.3.4.4	Defining the nucleus	142
4.3.4.5	Defining application programs and interface data formats (window)	142
4.3.4.6	Programming languages	142
4.3.5	Developing the nucleus	142
4.3.6	Initial database creation	143
4.3.7	Application program development	143
4.3.8	Documentation	144

