

COBOL PROGRAMMING

With

Business Applications

N. L. Sarda



PITAMBAR PUBLISHING COMPANY

COBOL PROGRAMMING

With
Business Applications

N. L. Sarda

*Computer Centre,
Indian Institute of Technology, Bombay.*



PITAMBAR PUBLISHING COMPANY

Educational Publishers

**888, East Park Road, Karol Bagh,
New Delhi-110005. (INDIA)**

Published by :

PITAMBAR PUBLISHING COMPANY

Educational Publishers,

888, East Park Road, Karol Bagh,

NEW DELHI-110005. (INDIA)

Telephones : { *Off. 519433, 526933* }
 { *Res. 562919, 561321* }

[First Edition : January 1979]

[Second Edition : January 1980]

[Copyright Reserved]

[Price : Paper Bound : Rs. 25/-

Hard Bound : Rs. 30/-

In Foreign

Countries : U.S. \$ Five only]

Printed at :

PITAMBAR PRINTING PRESS

41 & 42, Basti Harphoolsingh,

DELHI-110006.

[Telephone : 512041]

PREFACE

COBOL was designed around 1959-60 with a purpose to provide a common language for business applications of computers. It is now the most widely used language for data processing. To maintain uniformity and to extend the facilities, the Conference on Data Systems Languages (CODASYL) makes recommendations, which are endorsed by the American National Standard Institute (ANSI) and accepted all over the world. This book presents COBOL as per the 1974 standards. Except for a few highly specialized facilities like network communication and COPY, this book covers all features of COBOL.

This book has evolved from the extensive notes used by the author over the last five years for the courses on COBOL Programming and Automatic Data Processing at IIT Bombay and at some industries in Bombay. The contents have largely been referred by students and trainees of differing backgrounds, and, after some initial experimentation, the present layout of the course material has been found most satisfactory.

The book not only details rules and conventions of COBOL, but also justifies the data processing framework integrated into COBOL. The requirements of business applications, special nature of problems, and implications of using certain methodologies have been clearly brought out. Hence, it should serve as text both for learning syntax and semantics of COBOL and for getting an insight into business problems and training in designing systems. No assumptions have been made about the background of readers, except for some awareness of business organisations and their functions.

COBOL is a tool, not the target. Hence, it is necessary to orient the text towards applications also. A number of complete, comprehensive, worked-out examples are included which not only bring out usage of COBOL but also impart a training in utilization of computers. Numerous, interesting, and challenging exercises are given at the end of each chapter to test the reader in both fundamentals and practice. For the texts of this kind, it is always a source of difficulty to decide on nature of applications to be selected for examples and exercises. To avoid too much discussion on business activities and thereby divert from the main goal, we have selected examples and exercises from commonly familiar business situations. In order to thoroughly exercise the students, a list of project problems is included in Chapter 14.

It is indeed a great pleasure to acknowledge the encouragement and help I have received from my colleagues and friends. Specifically, I would like to record my thanks to Professor J. R. Issac and Mr. V.V. Kenetkar. I am also grateful to Professor S. Narasimhan, Co-ordinator of the Curriculum Development Centre at IIT Bombay for providing financial support and encouragement throughout the preparation of the manuscript. And to my wife and daughter—well, they understand my inability to express my gratitude to them!

—N.L. Sarda

ACKNOWLEDGEMENT

The following paragraphs are reproduced from the American National Standards Institute COBOL language standard X3.28.

“Any organization interested in using the COBOL specifications as the basis for an instruction manual or for any other purpose is free to do so. However, all such organizations are requested to reproduce this section as part of the introduction to the document. Those using a short passage, as in a book review, are requested to mention ‘COBOL’ in acknowledgement of the source, but need not quote this entire section.

COBOL is an industry language and is not property of any company or group of companies, or of any organization or group of organizations.

No warranty, express or implied, is made by any contributor or by the COBOL committee as to the accuracy and functioning of the programming system and language. Moreover, no responsibility is assumed by any contributor, or by the committee, in connection therewith.

Procedures have been established for the maintenance of COBOL. Inquiries concerning the procedures for proposing changes should be directed to the Executive Committee of the Conference on Data Systems Languages.

The authors and copyright holders of the copyrighted material used herein

FLOW-MATIC (Trademark of Sperry Rand Corporation),

Programming for the UNIVAC I and II, Data Automation

Systems copyrighted 1958, 1959, by Sperry Rand

Corporation ; IBM Commercial Translator Form No.

F28-8013, Copyrighted 1959 by IBM ; FACT, DSI

27A5260-2760, Copyrighted 1960 by Minneapolis-Honeywell.

have specifically authorized the use of this material in whole or in part, in the COBOL specifications. Such authorization extends to the reproduction and use of COBOL specifications in programming manuals as similar publications”.

FOR
My Mother,
Wife and Daughter

CONTENTS

Chapter

Preface

1. INTRODUCTION TO BUSINESS DATA PROCESSING 1

- 1.1 Nature of Business Requirements 1
- 1.2 Data Management Functions 3
 - Need for Improved Methods of Data Processing.*
- 1.3 Computer-based Data Processing Systems 4
- 1.4 Example : Branch Reconciliation for Banks 6
- Summary 10
- Bibliography 10
- Exercises 10

2. COMPUTER SYSTEMS 12

- 2.1 Introduction 12
- 2.2 The Users' View 12
- 2.3 The Physical View 15
- 2.4 Input Output Devices 16
 - General Characteristics, Punched-card Devices, Printer Devices, Magnetic Tape Devices, Magnetic Disk Devices, Teletype Devices.*
- Summary 21
- Bibliography 21
- Exercises 21

DATA AND ALGORITHMS 23

- 3.1 Introduction 23
- 3.2 Data Characteristics 23
 - Elementary Data Items, Arrays : An Introduction, Groups or Records, Files.*
- 3.3 Algorithms and their Representations 25
 - Algorithm, Algorithm Representation, Flow-charts, Decision Tables, Textual or Step-by-step Description, Representation in COBOL*
- 3.4 Examples 31
 - Preparing Bill for Purchases, Searching a Table, Internal Sorting, Summary of Sales.*
- Summary 34
- Bibliography 34
- Exercises 34

4. ELEMENTS OF COBOL 37

- 4.1 Introduction 37
- 4.2 Structure of COBOL Programs 37
- 4.3 Preliminary Features of COBOL 38

- 4.4 Layout of COBOL Programs 40
- 4.5 General Punctuation Rules 40
- 4.6 Formal Method of Specifying COBOL Grammar Rules 42
- 4.7 The Identification and Environment Divisions 43
- Summary 44
- Exercises 44

5. PROCEDURE DIVISION STATEMENTS 46

- 5.1 Arithmetic Operations 46
 - Common Options, The ADD and SUBTRACT Statements, The MULTIPLY and DIVIDE Statements, Arithmetic Expressions and COMPUTE Statement.*
- 5.2 Data Transfer Statements 49
 - MOVE Statement, ACCEPT and DISPLAY Statements, Statements for File Processing.*
- 5.3 Statements for Control of Execution Flow 51
 - Sections and Paragraphs, STOP and GO TO Statements. IF Statement, Simple Conditions, Compound Conditions.*
- 5.4 Examples 54
 - Validation of Data, Processing of Quotations*
- Summary 56
- Exercises 56

6. DATA DESCRIPTION 58

- 6.1 Introduction 58
- 6.2 Elementary Data Items 59
 - PICTURE Clause, USAGE Clause, Editing and Edit Pictures, Padding, Truncation and Size Error.*
- 6.3 Additional Clauses for Elementary Items 63
- 6.4 Groups and Tree Structures 64
- 6.5 Arrays and OCCURS Clause 67
 - Higher Dimensional Arrays, Subscripts, Arrays of Varying Size, Rules to Remember.*
- 6.6 Other Useful Facilities 70
 - Qualification, Condition Names Level 77 : Independent Elementary Items, The data-name FILLER, The REDEFINES and RENAMES Clauses.*
- 6.7 On Design of Files and Records 73
- 6.8 Examples 75
 - Validation Program, Sales Summary, A Patient Data Base.*
- Summary 78
- Exercises 78

7. FILE CHARACTERISTICS AND RELATED COBOL SPECIFICATIONS 82

- 7.1 Introduction 82
- 7.2 General File Characteristics 83
 - File Function, File and File-record, File-record and its Contents, File and File Volume, File and Device Records : Blocking and Deblocking, Buffering for Sequential Files, File Protection, Errors in File Processing, Other Miscellaneous Characteristics.*
- 7.3 File Organization and Access Mode 90
 - Sequential File Organization, Indexed File Organization, Relative File Organization.*
- 7.4 SELECT and FD Entry Formats 94

- 7.5 Comparison of File Organization Methods 95
- 7.6 On Selection of File Organization 97
- 7.7 Examples 98
 - Copying from Cards to Tape : 4 Cases, Blocking on Cardfile.*
 - Summary 102
 - Exercises 102
- 8. SEQUENTIAL FILE PROCESSING 104
 - 8.1 Methodology of Sequential Data Processing 104
 - Batching of Transactions, File Maintenance, Update Procedure.*
 - 8.2 COBOL Statements for Sequential Processing 108
 - General Formats, Programming for Printer Files and Special Options*
 - 8.3 Examples 110
 - Sales Summary Report, A Payroll System.*
 - Summary 115
 - Bibliography 115
 - Exercises 115
- 9. INDEXED AND RELATIVE FILE PROCESSING 119
 - 9.1 Introduction 119
 - 9.2 Indexed File Processing 120
 - Creation, Sequential Retrieval, Random Retrieval, Updating.*
 - 9.3 Relative File Processing 122
 - Hashing and Collision Handling, Creation, Retrieval and Update.*
 - 9.4 Examples 125
 - Inventory Updating, Accounts System for a Bank.*
 - Summary 129
 - Bibliography 130
 - Exercises 130
- 10. ADVANCED PROCEDURE DIVISION FACILITIES 132
 - 10.1 The PERFORM Statement 132
 - Introduction, Subroutine Call Function, The EXIT Statement, Repetitive Execution Function.*
 - 10.2 Miscellaneous Statements and Options 137
 - CORRESPONDING Option, ALTER Statement, GO TO with DEPENDING Option, ACCEPT Option.*
 - 10.3 Indexing and Table Searching 139
 - Indexing, SET Statements, SEARCH Statements.*
 - 10.4 The String Processing Statements 143
 - STRING Statements, UNSTRING Statement, INSPECT Statement.*
 - 10.5 Debugging Aids 146
 - 10.6 DECLARATIVES 147
 - 10.7 Inter-program Communication 148
 - Summary 149
 - Exercises 149
- 11. REPORT WRITER FACILITY 153
 - 11.1 General Characteristics of Reports 153
 - 11.2 Advantages of Report Writer 155

- 11.3 Page Formatting 156
- 11.4 Report Description 156
- 11.5 Statements for Report Writer 159
- 11.6 Example 160

Summary Payroll Report.

Summary 162

Exercises 162

12. SORT/MERGE FACILITY 165

12.1 Introduction 165

12.2 The Keys, their Significance Orders, and Sorting Orders 166

12.3 Input to and Output from SORT/MERGE Processes 166

12.4 Sort-file and Merge-file 167

12.5 The Related Statements 168

12.6 Example 169

Sales Contest.

Summary 171

Exercises 171

13. SOME ASPECTS OF ENVIRONMENT, DESIGN AND FUTURE OF DATA PROCESSING 173

13.1 Introduction 173

13.2 Environmental Aspects 173

13.3 System Analysis and Design 176

13.4 Data Processing in Future 179

Bibliography 182

14. PROJECT PROBLEMS 184

Appendix I : List of COBOL Reserved Words (i)

INDEX to terms and Concepts (iii)

INDEX to COBOL Syntax Rules (v)



INTRODUCTION TO BUSINESS DATA PROCESSING

1.1 NATURE OF BUSINESS REQUIREMENTS

We all have an intuitive understanding about the meanings of the terms 'data', 'processing' and 'information'. Consider a retail store and purchase of some commodities by a customer at the store. The cost of each commodity is calculated by multiplying its unit-price and the quantity-purchased, and the total cost is obtained by adding costs of all commodities. The transaction of purchase is thus described by data containing name, unit-price and quantity of each commodity purchased, the processing of data consists of multiplications and additions, and the information is the total-cost to be used in charging the customer. The purchase transaction, in general, is characterized by the attributes or properties called commodity name, price and quantity. The values of these properties for a specific purchase transaction constitute the data. For example, the values

pencil, 55 paise, 10 ^{the worth of} _{sth. in money}
are data describing one purchase-event—the purchase of 10 pencils.

The word 'datum', of which the word 'data' is plural, means a fact, or more specifically, a value of an attribute of an 'entity'. An entity, in general, may be an object, idea, event, condition, or situation. Thus, an employee, a product, a purchase-event

are entities. A set of attributes describes an entity. The values of these attributes are data about the entity. Usually, we select only those attributes of an entity which are relevant to us. For example, the colour of commodity purchased by the customer is not important in calculation of the cost. Hence, although colour is the property of commodity, it is not relevant. Similarly, for paying the salary to an employee, his height is not relevant. On the other hand, his name, basic salary, provident-fund contribution rate are relevant.

The term 'data processing' implies transformation of data for producing results useful in carrying out the activities of business. The result of data processing is called information. The transformations on data consist of a sequence of operations. The sequence is also called a 'procedure'. Obviously, a procedure depends on the nature of business and the kind of information required. A 'data processing system' is a system which takes data as input, carries out the required processing on data, and produces the information. The system is called 'manual' when the processing is performed by human beings, and 'automatic' when machines are used for processing. The situation is analogous to a manufacturing organization producing finished products from raw-material. Data are used as raw-material for producing information. The difference

between data and information is thus 'what exists' and 'what is derived'. Quite often, the two terms are used interchangeably, since there may be multiple levels of processing.

The organization or the business acts as an environment for its data processing system. The data originate in, and the information is consumed by, the environment. The information is used both for the purpose of day-to-day administration and policy decisions. Data processing is ubiquitous, whatever may be the nature and size of an organization.

Besides its use in producing information, data requires to be managed. For the retail store, the table of unit-prices must be so organized that price of an item can be easily located. The table must be updated as market prices change, and extended when new commodities are introduced for sale. It is also necessary to keep track of stocks-in-hand so that depletions do not come as surprise and affect the business, to keep a list of suppliers, to place purchase orders, and to maintain accounts of payments. The management of data for a retail store is not immediately apparent because it is on a small scale, and can be managed by storekeeper manually.

As the complexity of business increases, the number of functions to be performed and resources to be controlled also increase. A way of managing complexity is by dividing the task into subtasks. The business organizations have hierarchies of departments and sections, each with a well-defined area of responsibility and authority. A manufacturing organization may have production, procurement, sales, account, and personnel subdivisions. Besides having to perform specific functions, they must interact with each other in achieving the overall goals of the organization.

The complexity of data processing and management of data is proportional to that of the organization. A large number of different types of entities interact within the environment. Typically, the different types of entities are employees, products, machines, warehouses, sales-orders, suppliers, customers, production schedules, work orders etc. Similarly, various kinds of information are required for the different functions. Different types of data, describing the different types of entities, must be made available for processing as *transactions* take place within the organization. A transaction can be regarded

as interaction between two or more entities. Data originate as the result of interaction or activities of the entities, are processed in the subdivisions, and flow through the established channels in the hierarchy of the organization. For example, consider the activities resulting from a sales-order: an order from a customer is received by the sales department; the availability of the requested product is checked; a despatch notice is sent to a warehouse; an invoice is prepared and sent to the accounts department; the inventory is updated; a production order is sent to the production department; availability of raw-material is checked and may result in placing a purchase order for acquiring the raw-material; invoice is sent to the customer after receiving the acknowledgment of delivery; an entry is made in the accounts-receivable register; the customer account is updated; etc. The picture is very much different compared to the retail store, and we find that the transaction of a sale cannot be dispensed with immediately. Data must be recorded and processed to produce various kinds of information, referred back again, and communicated to other departments.

The processing and management of data is very crucial for any organization. Whatever may be the nature of organization, government or private, profit-making or non-profit-making, manufacturing or educational, data processing is as important to it as any of its other functions, or even more so. An organization may use resources like men, machines, energy, money, etc., but none of these can give a complete picture of the state of affairs. Data is the most important resource as it is the only one which can do so. In the past, there was a tendency to look upon data processing as a low-level, clerical task; however, the importance of processing and management of data became clear as the organizations became complex and dynamic, as the principles of decision-making were understood, and as the realization grew that the coordination of different activities of the organization is essential for efficient service and use of resources. Data, in a sense, being the past history of an organization, contains information that can be used both in effective administration and planning. The importance of data processing is aptly reflected in the new name it has acquired: 'Information Systems'.

1.2 DATA MANAGEMENT FUNCTIONS

Data usually originate on documents. In a manual data processing systems, the clerical or administrative staff collect, classify, store in files, and use the documents. These functions must be performed even in an automatic data processing system. Although the details may differ from one organization to another, the basic functions performed in managing the data are same. The various functions are discussed below :

(i) *Collection.* The data originate at a specific source in the organization. The different types of data originate at different sources, which may be geographically distributed. The documents containing data must be collected at the sources and passed on to the data processing system.

(ii) *Classification.* The documents must be classified into groups depending on nature and use of data contained in them. The classification is necessary in order to facilitate systematic storage of and easy reference to the documents. It immediately sounds absurd to store together bio-data of employees and addresses of suppliers of raw-materials, or marks obtained by students and monthly earnings of teachers. Classification essentially groups entities of similar type ; for examples, the groups may be all employees, all suppliers, all students, etc. The entities within each group have similar properties and are used in similar way in the processing of data. Data corresponding to one group of entities is stored as a file. The term 'file', in a general sense, means a collection of data. A major classification of data automatically takes place when an organization is divided into subdivisions, each having its own data processing system.

(iii) *Sorting.* The physical ordering of data within a class, stored as a file, is called sorting. Data about different entities are arranged in a specific order, based on one or more attributes of the entities. For example, every supplier has a name ; arranging the supplier data in alphabetical order of supplier-names is sorting. It makes the search for a particular supplier within the file much easier compared to when the file is in random order. The sorting of data is a very common operation in business data processing. In many cases, re-ordering of data itself acts as information. The supplier data may be sorted by prices quoted by them in selecting a supplier

for placing an order, or they may be sorted by delivery period when it is necessary to obtain supplies in shortest time. Sorting is always with respect to the values of certain property of the entities. The supplier-name, price and delivery period are all properties of supplier-entities. The order of sorting is called ascending or descending depending on whether the data is arranged in increasing or decreasing order of the values of the selected property.

(iv) *Storing.* It is the most obvious function in data management. Data must be stored in order to refer to it in future. The paper documents are held together in a file, the files are stored into compartments of a cabinet, etc. The paper acts as a medium for storage of data. An automatic system may use different media for data storage. Not only the data within a file, but also the files themselves must be stored in some systematic manner. Data are stored after their classification and sorting for the reasons given above. The physical way of storing data is also called as 'data organization'. The collection of all data is also known as 'data base'. Sometimes, additional data is prepared to facilitate searching for specific data in the data base. The book-index, at the end and the contents at the beginning of a book are typical examples of the additional data.

(v) *Retrieval.* The process of locating and extracting required data from a data base is called 'retrieval'. The manner of retrieving data is called 'method of accessing' or 'access mode'. There are basically two methods of data accessing and their use depend on the organization of data or the purpose of retrieval. Let us consider a simple example to illustrate the two methods of accessing. A file containing a list of supplier-addresses is given and it is required to send a request for quotation (RFQ) to each of them. Starting from the beginning of file, taking one address at a time, we copy the address on RFQ. That is, we proceed in the same sequence in which the addresses appear in the file, and retrieve every address. This method of accessing is known as 'sequential'. On the other hand, assume that a purchase order is to be placed with a specific supplier, whose address must be obtained from the file. The name of the supplier is known. (Let us assume that all suppliers have unique names.) If the addresses are stored in random manner (i.e., due to the random organization of data), we are again forced to scan the file in

sequential access mode until the desired address is found. However, if the data is systematically organized or an index by supplier-name is available, it is possible to directly locate the desired address. In the latter case, the access mode is called 'random', implying that the desired data can be located randomly as the need arises, without going through irrelevant data. Note that in random access mode, data is accessed using a property of the corresponding entity. In the above example, name of the supplier was known. Also note that the organization of data must facilitate random access, and if it does not we must use sequential access mode.

(vi) *Maintenance*. The objective of this function is to keep the stored data up-to-date, or in a state reflecting the current status of organization. It involves making changes to, deletions from, or additions to the existing data. How up-to-date or current the stored data should be depends on the application, and has a great bearing on the method of data processing. The employee data may be updated once in a month (just before making the monthly payments), the stock (inventory) updates may be weekly, and accounts receivable updates as soon as payments are received from customers.

(vii) *Verification/Validation*. The data entering the system must be checked for accuracy. The term 'verification' means ascertaining that recording of data on paper or any other medium is correct, while 'validation' means making sure that recording of data is as per the prescribed rules. For example, ascertaining that the total-cost is sum of individual commodity costs is verification, while ascertaining that items and quantities are according to the customer request is validation. In practice, this fine distinction between the use of these two terms is not maintained.

Besides these important functions, the following functions are also used as and when the need arises :

(viii) *Transcription*. Changing the medium on which data is recorded. It is a common function when machines are used for data processing, since machines usually do not accept hand-written or typed documents on which data is initially recorded.

(ix) *Duplication*. Making multiple copies of data.

(x) *Coding*. Assigning unique codes to

entities for their identification within a class or group of similar entities. A code may also indicate certain properties of the entities. For example, a salesman may be identified by a 4-digit code, of which the first digit indicates the region to which he is attached. A code must both be concise and precise.

(xi) *Editing*. Putting the data in more presentable or useful form.

(xii) *Collecting/merging*. Combining two sets of similarly sorted data.

(xiii) *Transmission*. Transmission or communication of data or information from one place to another.

NEED FOR IMPROVED METHODS OF DATA PROCESSING

The management of an organization is responsible both for administration and planning, and for either of these functions it needs information on which to base its decisions. The necessary data processing may be carried out manually or by using machines. The organization operates continuously and as such it must be continuously monitored. The data processing system must be responsive enough to supply the information when it is needed, so that the performance of the organization and effectiveness of the decisions can be measured and corrective actions can be taken. The value of information reduces with time, and the data processing system is not considered satisfactory if it cannot supply the information within time. Large organizations must handle large volume of different types of data. The manual methods of data processing often fail due to long processing times, inconsistency, and unreliability. Moreover, ad hoc information needs of the management cannot be immediately satisfied. Sometimes, the clerical costs of processing the data are exorbitant and forces the management to do without the processing of data.

To successfully meet the information requirements of the modern, complex, highly-dynamic business organizations, it is necessary to make use of machines for data processing. The machines have advantages of speed, consistency, and reliability over human beings. Since their introduction in fifties, the computers have established themselves as a valuable automatic tool for data processing.

1.3 COMPUTER-BASED DATA PROCESSING SYSTEMS

A computer differs from other types of

machines in that it is the only one where transformation of energy is not an objective; it is a machine designed for transformation of data. It can accept data, carry out arithmetic/logical operations and tests on data, and output the results of manipulation. Moreover, a computer is a 'general-purpose' machine: it can be 'programmed' to achieve desired transformations on any given data. A computer can prepare payrolls as well as invoices, plan production schedules and maintain inventories. The only requirement of the task for computerization is that the processing within the task be expressable in unambiguous form and using operations that can be performed by the computer. A computer can not handle unexpected situations and cannot make non-predetermined decisions. Programming a computer consists of describing the characteristics of input and output data, and specifying the procedure containing sequence of operations to be performed on the data. It is thus a tool which can both supplement and complement information processing capability of a human.

The advantages of using a computer for processing of data were mentioned above. Whatever may be the tool used in data processing, the gross nature of and functions within the data processing system are same as described earlier. The data must be collected, classified, sorted, stored, retrieved and maintained; procedures must be defined and 'executed' to obtain results and to maintain the data base. The purpose of this section is to give a preliminary introduction to similarities and differences in some of the important aspects of manual and computer data processing.

While a human being uses his intelligence and experience and can interpret instructions given to him from their context, a computer must be instructed in precise terms and using only those operations which are built into it. For a computer, data is just a string of symbols, and it is the user's responsibility to manipulate data appropriately. The meaning of data or outputs from computer must be in the mind of the user. For example, the two expressions :

unit-price X number-of-items, and
daily-wage X number-of-days-worked

are very much similar for a computer. The differentiation must be through our interpretation. While the training and experience of a human being can be used in instructing

him, and we can comfortably ask him questions like 'how many employees get salary more than Rs. 800', a computer must be instructed about the procedure for obtaining answers to such questions. The data must be supplied in the format expected by the computer (i.e., as per its instructions), and even a slight deviation, so trivial as never to be mis-interpreted by a human, may result in rejection of data, or even worse, may be wrongly used.

It is the responsibility of users to assure correctness of data and procedures supplied to a computer. The steps in a procedure are followed by the computer in what we may call a blind manner. The procedure should include steps for checking reasonable accuracy of data.

The source data, as said before, usually originate on documents. Unlike humans, a computer cannot read them directly. The relevant data from the documents must be transcribed onto a medium acceptable to the computer. In Chapter 2, we will consider media used by computers. Transcription is also known as 'data preparation'. Special devices are available to perform transcription. The results from the computer can be obtained in printed form, requiring no transcription at the output end. It is an added responsibility of user to verify that no serious errors are made during the transcription.

The computer ^{uses} media which permit high density of recording, permitting storage of large volume of data in very small space, thus reducing storage and handling problems. It also uses very high rates of data transfer and very high computation speeds, which enable it to carry out data management and data processing tasks on a scale unimaginable for humans. Thus, what used to be considered as too expensive in cost, effort and time for clerical staff can be effectively realised with a computer.

A computer-based data processing system has four components corresponding to its manual counterpart: files which contain data, programs which incorporate description of data and procedures, outputs which contain the required information, and a schedule for 'executing' the programs indicating the sequence in which the procedures must be performed.

The concept of a file is central to the management of data on a computer. The files contain data from earlier processing, or

data required in future processing, or both. Data are stored within files after proper classification: consequently, a file contains data about entities of similar type. As mentioned earlier, this is so also for files manually maintained. Data about one entity form a 'record'; thus, a file is also a collection of records. The records contain similar type of data about a class of entities. A property of an entity appears as a 'field' or 'data-item' in the record. We can indicate the hierarchy of existence of data in reality and in a data processing system, and their mutual correspondence as follows:

(i) In real world:

entity class — entity — property of entity — value of the property

(ii) In data processing system:

file — record — field or data item — string of symbols

(iii) An example:

all employees — one employee — salary — 1247.35

Data which originate in the organization as a result of activities or events are commonly called as transactions. They are transcribed and processed along with other files. They may cause the programs to update data within the existing files. File is said to be 'off-line' if it is not accessible to the computer, and it becomes 'on-line' when mounted on a device controlled by the computer. As with manual data processing system, the transactions may be processed as soon as they originate, or they may be accumulated over a period of time and processed together. In the former case, the data processing system is called 'on-line' and in the latter case as 'batch' processing system. In both the manual and automatic systems, the physical organization of data is greatly influenced by the way of processing transactions.

Fig. 1.1 shows the major steps of a computer-based payroll system. Such a chart illustrating functions and sequence of steps is called a 'process chart'. The steps may be executed daily, weekly or monthly depending on the payment policy of the organization. Source documents may be employee time cards indicating hours worked on each day of the

period. The source data is first transcribed onto a medium acceptable to the computer, and then 'passed through' a program which checks for correctness of the data. The program embodies the testing of conditions which must be satisfied by the data. The erroneous data is rejected; the correct data may undergo further processing, or it may be delayed until corrections are made and received. There are two files which contain payroll data. File 1 contains data for calculating earnings from the work data, while file 2 contains the accumulated savings and tax-deductions for each employee. It is assumed that the data in files are sorted by employee codes. The work data is first sorted to bring it in the same order as the two files. Finally, using the work data and salary data the earnings, savings, and deductions are calculated. The data in file 2 is updated to include recent savings and deductions. The figure uses boxes of different shapes for different functions. A rectangle with a zig-zag bottom stands for data/information on paper medium, a rhombus to indicate a manual activity, a rectangle for computer processing, and a circle for a file. The arrows from a file indicate direction of data transfer.

The payroll system discussed above is an example of batch processing. Note the preciseness of each step. A manual system may function exactly as in Fig. 1.1, or it may completely process data of one employee at a time. Once they are defined, the steps of the process-chart are rigidly followed in computer data processing.

1.4 EXAMPLE: BRANCH RECONCILIATION FOR BANKS

In Sec. 1.1, we pointed out that management of data is an important function in data processing. The purpose of this section is to illustrate that the procedures of business data processing are usually simple, that they are highly repetitive in nature, and that need to manage large volume and different types of data is responsible for the complexity in business data processing.

A national bank has its branches spread all over the country. There are inter-branch transactions of various types like mail transfers, demand drafts, debit notes, credit notes, etc., each resulting in transfer of money from one branch to another. The sending branch classifies the transactions as debit since the transactions reduce its balance; the receiving

branch on the other hand classifies them as credit transactions. The objective of reconciliation is to match appropriate credit and debit entries made by branches in their own accounts. The procedure of data processing

is very simple : it just consists of comparing two transactions and verifying that they match. However, the simple question 'how to locate the two matching transactions, if they exist, in a collection of transactions'

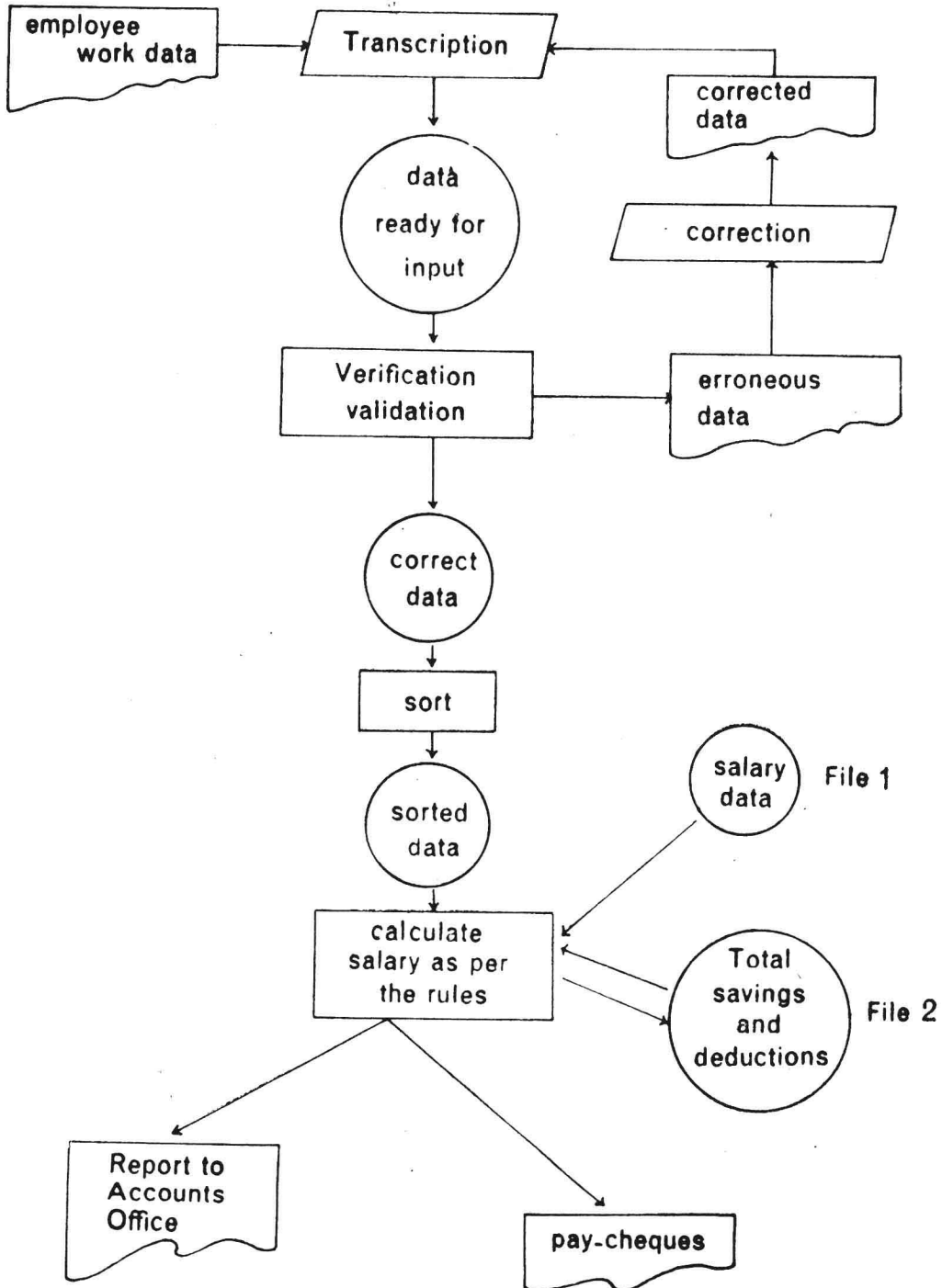


Fig. 1.1. : Steps in a payroll system. Circles indicate files.