ILLEGIBLE DOCUMENT

THE FOLLOWING
DOCUMENT(S) IS OF
POOR LEGIBILITY IN
THE ORIGINAL

THIS IS THE BEST
COPY AVAILABLE

ANALYSIS OF A CONVERSION TO
DISTRIBUTED DATA PROCESSING

by

JOSE GENARO SERGIO SILVA LOPEZ

B.S., Escuela Superior de Ingeniería Mecánica y Eléctrica (Mexico), 1970

_____

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1979

Approved by:

_____
Major Professor

LIST OF FIGURES

With love

to my wife, Elizabeth

and my children

Elizabeth, Sergio, and Erica

INDEX

INDEX (continued)

I. Background

This is a study of one bank which is a nationwide service organization, consisting of a central office and regional offices. Regional offices are distributed across the country according to geographical locations. Each regional office controls branches that are allocated in that region.

The bank presently runs its applications at central office on a large computer. Over the years both the bank's business and organization have expanded to the point where the current centralized system no longer has the ability to satisfy information needs within the organization; instead operations are becoming increasingly difficult to handle. Along with this, management has issued a general policy, indicating a strong desire for decentralization.

II. Introduction

A bank performs most of its operations based on the information contained on several data bases which constitute the information system of the bank. At present this information system is supported by a large computer in a centralized fashion. In the near past this type of centralized processing was good enough to support all the required transactions as well as decision making.

Now, new applications and a big increment in the number of transactions have been producing some problems in system operations. Along with this, management of the bank has issued a general policy, in which is explained management's desire for decentralization of all those operations that can be managed at regional offices without intervention of the central office. Likewise, management has recognized the possibility of decentralizing EDP operations in order to permit an easy and opportune flow of information as well as a good information control. However, management realizes the difficulties that arise when converting from centralized data base operations to distributed data base operations.

The purpose of this master's report is to present insights, requirements and conditions that may be taken into consideration when moving from centralized data base processing to distributed data base processing.

The bank's loan system is used as a sample in this master's report, in order to accomplish the above objectives.

The implementation of the loan system will provide the experience in order to evaluate and implement the decentralizatin of other systems. System characteristics that need to be taken into consideration for

system evaluation are data base size, system performance, user characteristics, backup system, etc. A detailed description of such characteristics is shown in the next chapter.

III.   Current System

Among the data base systems that are constituting the information
system of the bank, the bank's loan system has been selected because
it involves almost the same type of operations that are carried out
in other systems such as checking accounts or savings accounts. How-
ever, the loan system is selected for the initial step in decentrali-
zation because it is not a critical system. This is due to the low
frequency of transactions. Therefore the nature of the system will
permit us to avoid situations such as the pressure presented in other
systems in the event of hardware or software failure.

In order to describe the current system, three stages are going
to be presented. These stages are data capture, data preparation,
and consulting the data base.

3.1   Data Capture Description

As it was explained in chapter I, the bank studied is a nation-
wide organization consisting of a central office, central branches,
regional offices and regional branches (fig. III.1). Therefore, data
capture is based on regional locality, and the data source travels
from the branches to the central office according to the following
procedure (see fig. III.2).

3.1.1 Data Capture Procedure

a) Once the loan is approved, a form (see appendix A) is coded at
the branch that has granted such a loan. A copy of this document is
kept in an envelope, the grant is recorded on a list, and the branch
continues with its own operations.

DATA CAPTURE FLOW DIAGRAM

CENTRAL AREA

CENTRAL BRANCH   CENTRAL BRANCH   . . .   CENTRAL BRANCH

C E N T R A L   O F F I C E

REGIONAL OFFICE   . . .   REGIONAL OFFICE

RB   RB   . . .   RB   RB   RB   RB

OFFICES DISTRIBUTION BY REGIONS

RB   Regional Branch
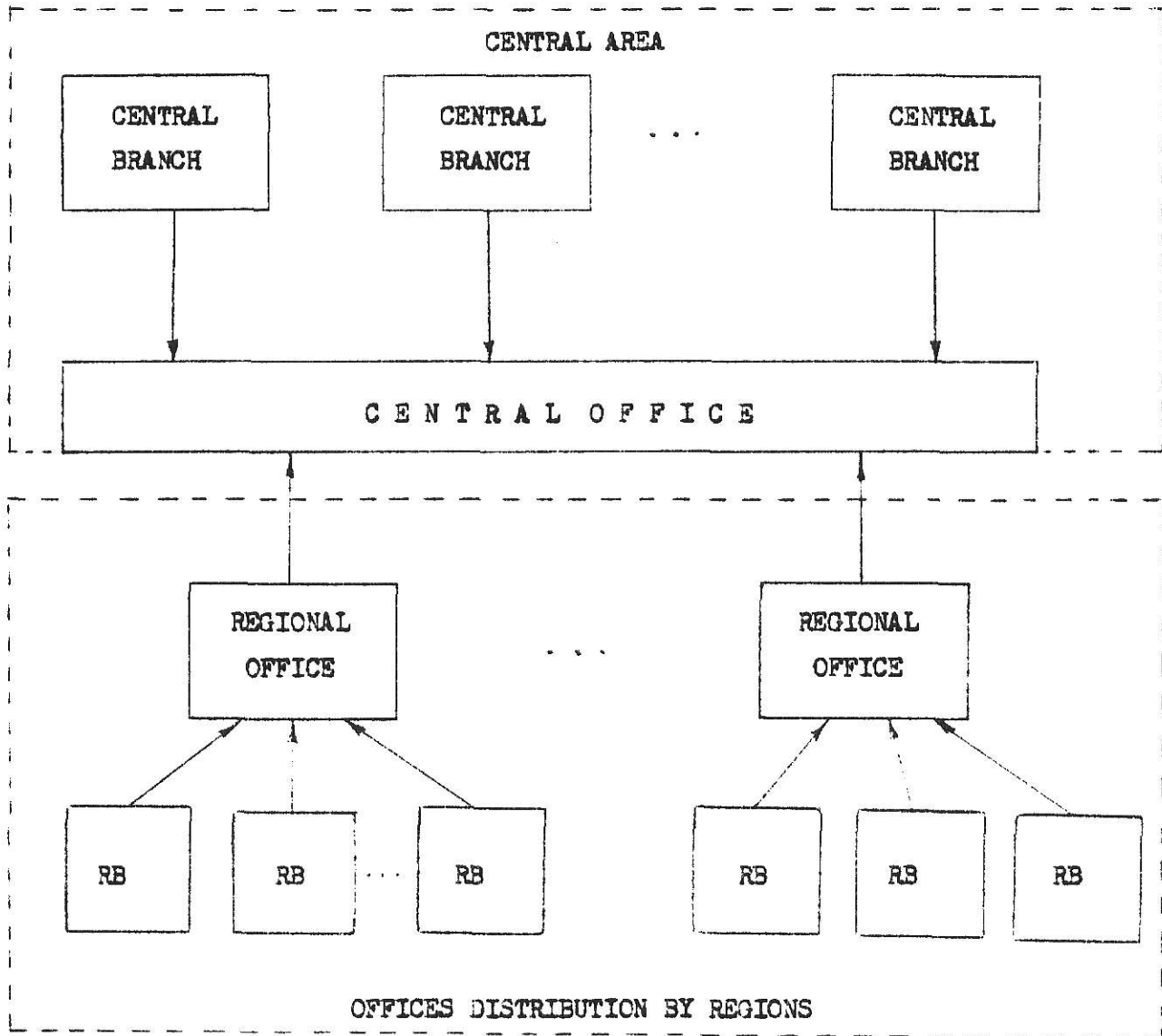
⟶   Data Source Flow

(Fig. III.1)

DATA CAPTURE PROCEDURE

```
                    ┌─────────────┐
                    │    START    │
                    └─────────────┘
                           │
        ┌──────────────────┤
        │                  ▼
        │              ╱ IS ╲
        │             ╱ LOAN ╲      NO      ┌─────────────┐
        │            ╱ APPROVED?╲──────────▶│  INFORM TO  │
        │            ╲          ╱           │   PERSON    │
        │             ╲        ╱            └─────────────┘
        │              ╲      ╱                    │
        │                 │ YES                    ▼
        │                 ▼                 ╭─────────────╮
        │          ┌─────────────┐          │   END OF    │
        │          │   CODING    │          │  OPERATION  │
        │          │    FORM      │          ╰─────────────╯
        │          └─────────────┘
        │                 │
        │                 ▼
        │          ┌─────────────┐
        │          │  KEEP FORM  │
        │          │ IN ENVELOPE │
        │          └─────────────┘
        │                 │
        │                 ▼
        │          ┌─────────────┐
        │          │  RECORDING  │
        │          │   ON LIST   │
        │          └─────────────┘
        │                 │
        │                 ▼
        │    NO        ╱ LAST ╲
        └────────────╱ WORKING ╲
                     ╲  DAY?   ╱
                      ╲       ╱
                         │ YES
                         ▼
                       ( A )
```

(Fig. III.2)

```
                          ( A )
                           │
                           ▼
      NO              ╱ CENTRAL ╲              YES
    ┌─────────────────   AREA ?  ───────────────┐
    │                  ╲         ╱               │
    │                   ╲       ╱                │
    ▼                                            ▼
┌──────────────┐                          ┌──────────────┐
│   BRANCH     │                          │   BRANCH     │
│ SENDS ENVE-  │                          │ SENDS ENVE-  │
│   LOPE TO    │                          │   LOPE TO    │
│  REGIONAL    │                          │  CENTRAL     │
│   BRANCH     │                          │   OFFICE     │
└──────────────┘                          └──────────────┘
    │                                            │
    ▼                                            │
┌──────────────┐                                 │
│  REGIONAL    │                                 │
│ BRANCH COL-  │                                 │
│  LECTS ALL   │                                 │
│  ENVELOPES   │                                 │
│  IN ITS AREA │                                 │
└──────────────┘                                 │
    │                                            │
    ▼                                            │
┌──────────────┐                                 │
│ SENDS ENVE-  │                                 │
│  LOPES TO    │                                 │
│   CENTRAL    │                                 │
│   OFFICE     │                                 │
└──────────────┘                                 │
    │                                            │
    │        ┌──────────────────┐                │
    │        │  CENTRAL OF-     │                │
    └───────▶│ FICE RECEIVES    │◀───────────────┘
             │ AND CHECKS       │
             │ DOCUMENTS        │
             └──────────────────┘
                      │
                      ▼
             ┌──────────────────┐
             │     DATA         │
             │  PREPARATION     │
             │     STAGE        │
             └──────────────────┘
```

(Fig. III.2) continued

b) The last working day of every month, the documents contained
in the envelope are checked for completeness. If the branch is lo-
cated in the central area, then it sends its envelope to the central
office. If the branch is located in a region, then it sends its enve-
lope to the corresponding regional office. Once the regional office
has collected all envelopes from its branches, it sends those envelopes
to the central office. The transportation media is bus or plane
depending on the distance to be traveled.

c) The central office receives and checks all the documents.
Then that information is turned to data preparation stage.

3.2   Data Preparation Stage

3.2.1 General Description

In the central area most branches have been using devices such
as key punch, key-to-tape, or key-to-disk (depending on the place's
importance and number of transactions) in order to perform data trans-
cription from the coding loan form (Appendix A) to a physical data
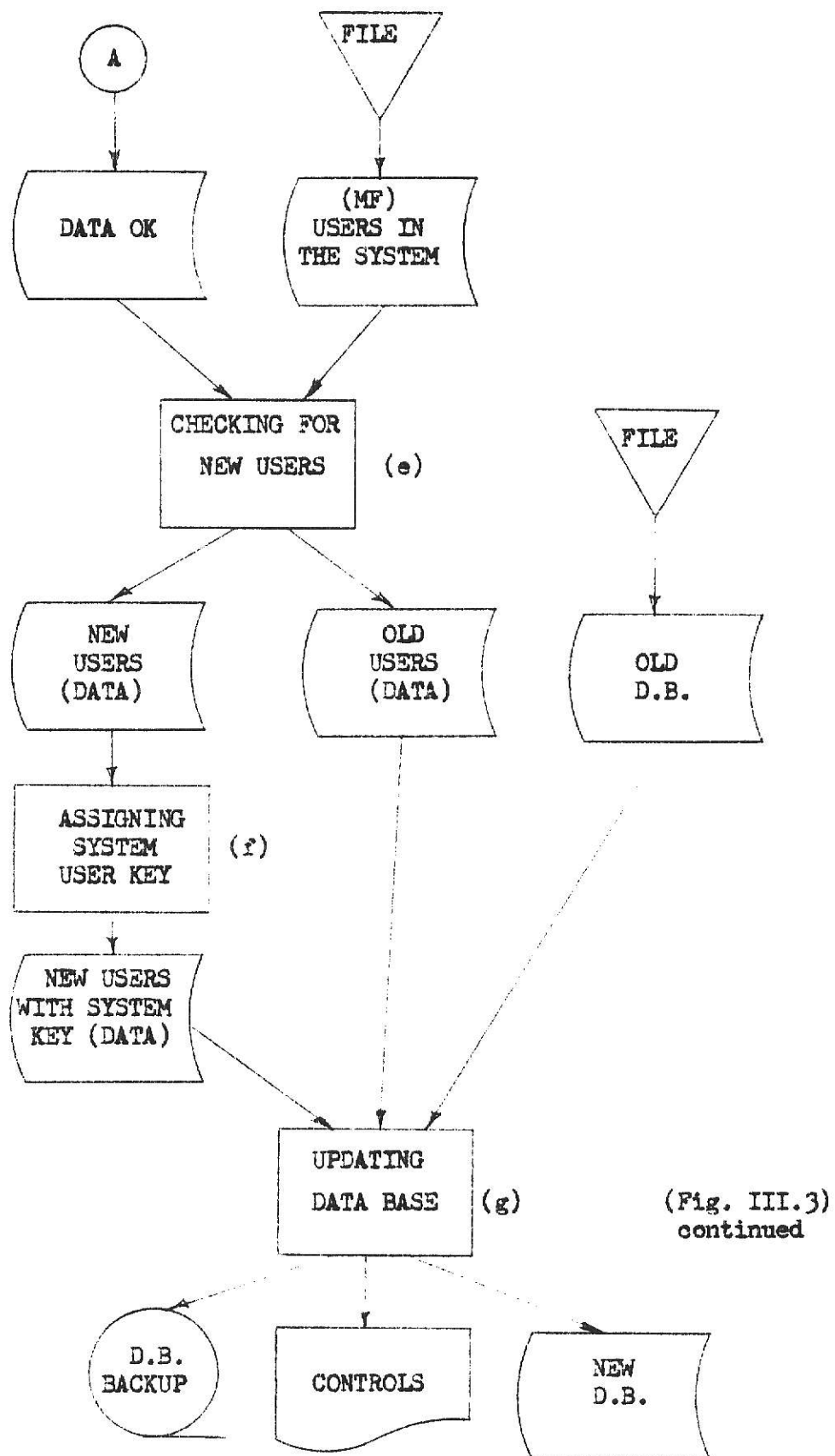support such as cards, tape or disk.

Many branches are using similar devices for data transcription,
thus all of them are producing the same type of code. However, a few
other branches are using rather older peripheral devices and they are
producing a different type of code.

At present all the data preparation stage is performed at the
central office. This stage starts as soon as the coded loan forms or
physical data supports arrive at the central office, and it is finished
when the data base is obtained and it is ready to be consulted (see fig.
III.3).

(3)

DATA PREPARATION PROCESS DIAGRAM

CLF = Coded Loan Form

```
                    ┌──────────┐
                    │  START   │
                    └────┬─────┘
                         │
                         ▼
            ┌─────────────────────┐
            │   RECEPTION OF      │
    (a)     │   SOURCE DATA AT    │
            │   CENTRAL OFFICE    │
            └─────────────────────┘
```

```
   CLF          SOURCE        SOURCE         SOURCE
 SOURCE          DATA          DATA           DATA
  DATA
```

```
   ▼
 ┌──────────┐                ┌─────────────────┐
 │   KEY    │   (b)   (c)    │     DATA        │  (OPTIONAL)
 │ PUNCHING │                │  TRANSLATION    │
 └──────────┘                └─────────────────┘
      │                             │
      ▼                             ▼
  SOURCE                      TRANSLATED
   DATA                       DATA SOURCE
```

```
            ┌──────────────┐
            │    DATA       │   (d)
            │  VALIDATION   │
            └──────────────┘
```

```
  ERRORS                       DATA OK
     │                            │
     ▼                            ▼
 ┌──────────┐                   ┌───┐
 │  MANUAL  │                   │ A │
 │CORRECTION│                   └───┘
 │ PROCESS  │
 └──────────┘
```

(Fig. III.3)

(9)

A

FILE

DATA OK

(MF)
USERS IN
THE SYSTEM

CHECKING FOR
NEW USERS   (e)

FILE

NEW
USERS
(DATA)

OLD
USERS
(DATA)

OLD
D.B.

ASSIGNING
SYSTEM
USER KEY   (f)

NEW USERS
WITH SYSTEM
KEY (DATA)

UPDATING
DATA BASE   (g)

D.B.
BACKUP

CONTROLS

NEW
D.B.

(10)

3.2.2 Detailed Description

Data preparation stage is constituted by seven possible processes which are described as follows.

a) Reception of source data. The central office receives at its Reception and Distribution department all the coded loan forms as well as the physical data supports.

b) Key punching process. The coded loan forms are sent to be processed at the key punching department in order to obtain the source data in a machine readable form.

c) Data translation process. This process is optional because not all branches are producing the same code. When the physical data support contains a different code than the common (ASCII), then the Reception and Distribution department fills in a form in order to request that the Operations department immediately translate the code.

d) Data validation process. This process accepts as input data the information coded in ASCII code. This input data can be contained on cards, tape or disk. The process is carried out by a program that accepts the input data, validates that data based on master tables and produces as output a formated listing presenting all errors contained on records. Also the program produces a file on disk containing all the information that has been considered acceptable.

The errors in the listing are corrected (*) and then the listing is sent to the punching department in order to obtain cards with correct

_____

(*) Sometimes the correction process requires a large number of phone calls, therefore the process can be slowed down considerably.

information. These corrected cards are used as input in the validation process. This correction process is repeated until the listing does not present errors.

e) Checking for new credit users process. Since source data can contain records with information of users already in the system and users that just received a loan, it is required to separate those two kinds of data. This process performs this task and produces one file containing data corresponding to new users in the system and another file containing data corresponding to old users in the system.

f) Assigning system key process. Each user must have a unique item identifier into his record. This process performs the task of assigning a unique key in the system to every new user. It is important to point out that one user may have received loans from different branches, however that user must have the same item identifier (key) in his records.

g) Updating data base process. Once the new users file and the old users file are obtained, these files are used as input to update the data base. As a result of this process the data base is obtained and is ready to be consulted. This data base is very different in its content, because of the deletions, insertions and modifications of records. Also, a listing is obtained showing the records that were deleted as well as the new number of records contained in each are of the scheme. Finally a backup of the data base is generated in order to be able to rebuild the data base in case of hardware or software failure.

All data preparation process is carried out in batch mode, and it is done once a month.

(12)

3.3   Consulting the Data Base.

3.3.1   Procedure to Access the Data Base.

All reference to the information contained in the data base are generated at the branches. Each branch requests information by telephone. If the branch is located in the central area, then the telephone call is directed to the central office. If the branch is located in a region then the telephone call is directed to its corresponding regional office. All phone calls are authenticated at the regional offices as well as at the central office. Each regional office records and keeps all queries that it receives. According to a schedule each regional office transmits all its queries to the central office. At the central office the consulting department receives all the queries and proceeds to type on a terminal the required commands in order to satisfy the queries. The answer to each query is sent back to each corresponding branch by telephone and also is mailed as hard copy (see fig. III.4).

As it is seen regional offices are operating only as query collectors, and according to a scheduling policy those queries are released to be answered at the central office.

3.4   Description of Problems on Current System.

The following problems in the current system have been identified. First of all, source data travels long distances in order to be processed (e.g. from regional branches to central office). Therefore, the data capture process is weak in terms of system integrity; not only because source documents can be lost when they are transported, but also when they are handled at the branch and/or at the central

CONSULTING PROCESS DIAGRAM

```
        Answer                    ┌──────────────┐
  ┌──────────────────────────────►│   BRANCHES   │
  ┆                               └──────┬───────┘
  ┆                                      │
  ┆              QUERY          telephone │
  ┆                                      ▼
┌─────────────────┐          ◇─────────────◇
│ COLLECT QUERIES │   YES    ╱   REGIONAL    ╲
│ UNTIL SOMETIME  │◄─────────   OFFICE        
│   ACCORDING     │          ╲               ╱
│   SCHEDULE      │            ◇─────────────◇
└────────┬────────┘                   │ NO
         │                            ▼
┌────────▼────────┐  phone   ┌──────────────────┐
│    REGIONAL     │          │  CENTRAL OFFICE   │
│ OFFICE TRANSMITS│          │ (Consulting Dept) │
│  ALL QUERIES    │   line   │ RECEIVES QUERIES  │
│                 │          │   AND TYPES       │
└─────────────────┘          │   COMMANDS        │
                             └─────────┬─────────┘
                                       │
        telephone        ╱──────╲      ▼
  ◄─────────────────────  DEMAND  ◄──── QUERY ────►┌───────────┐
                         │ MODE   │◄─── Answer ────│ PROCESSOR │
                          ╲──────╱                 └─────┬─────┘
                              │                          │
                              ▼                      ┌───▼───┐
                         ┌─────────┐                 │ DATA  │
                         │         │                 │       │
                         │         │                 │ BASE  │
                         └─────────┘                 └───────┘
                            Mail
```
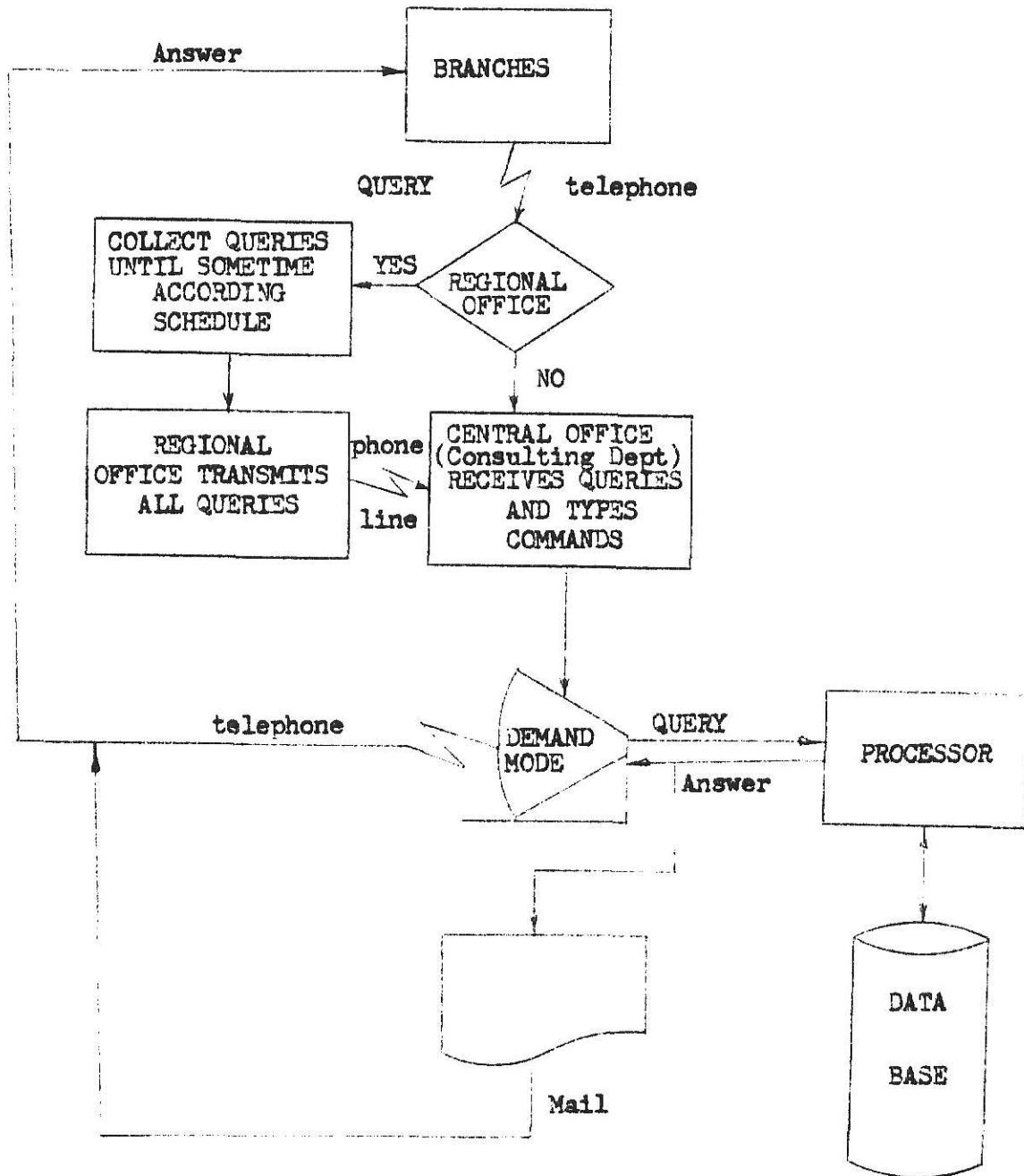
(Fig. III.4)

(14)

office. Also, sometimes the source information is received late at the central office because of transportation problems.

A bottleneck is produced at the central office when the source documents have been received and they are ready to be processed. In addition to this bottleneck, there is an information control problem. The central office can receive source information from branches on either coded loan forms, cards, tapes or disks at different times during the established reception period. Sometimes, the same group of data is sent to be processed more than one time. This kind of mistake also violates system integrity and at the same time slows down the data preparation process.

After validation process, a listing is obtained. This listing may contain errors which must be corrected. The correction process sometimes involves human actions. Since the information is processed at the central office and source data is generated at branches, then extra communication is required among people at central office and people at branches in order to clarify the data at central office. This extra communication slows down the process and also increases system costs.

Another problem arises at the consulting process. Queries from regional offices are transmitted to central office based on a scheduling policy (*). However, sometimes a large number of queries require a longer time than scheduled. Therefore, central office has to inform

_____

(*) Scheduling Policy indicates starting time at which each regional office can transmit its queries to central office as well as the time that can be used for this purpose.

(15)

every regional office of a new possible time to start the transmission (phone call). This situation produces another problem, because now personnel have to wait in order to transmit or receive information. Since personnel have to wait after regular time to leave the job; these personnel have to receive payment for overtime. This situation also increases the system costs, and sometimes personnel become unhappy with these occurrences.

Generally a user may request one loan at most in a month. However another problem arises when a user requests a loan in several branches during the same month. In this case the data base management system is going to provide incomplete information. This is due to the sequencing of data processing carried out in order to update the data base.

Data base administration (DBA) function also represents a problem. At present this function is handled by one individual. The system documentation is not completed because the data base administrator is busy in system maintenance or looking at new implementations. Therefore, this individual is becoming more important and powerful, not only because of his knowledge but also because there are no back up people to substitute for him in case of illness or resignation.

Data integrity and data security of the data base is a lesser, controllable problem; because data base operations are centralized in one site and controlled by one person. This means that if system hardware fails and if some portion of the data base or the entire data base is damaged, then recovery of the damaged data is an easy task, since the entire data base was backed up on tape, so a simple copy command enables us to have the data base again ready to be consulted.

It is important to point out that the data base is updated in batch mode and at the same time the backup is generated.

On the other hand, security is controlled by the data base administrator, because he is the only one who handles the schema and subschemas of the data base. Also access to the data base is done (according to designed formats by the DBA) by clerical personnel who only need to know how to type queries on the terminals.

Finally, system resources are underutilized; this is because the software system is designed in order to handle every possible query from every branch. Therefore all resources have to be loaded even though not all of them are going to be utilized.

IV. New Distributed System Approach

It is important to point out that an analysis was carried out in order to choose a distributed system that is feasible and at the same time best fit into the current and future information needs of the bank. This study was based not only on the future data bases system's expansion but also on the management's agreement on financial support for decentralization.

Before we start with the description of the new system approach, it is worthwhile to make a brief description of two systems that represent the extremes of data processing systems.

4.1 Centralized Systems

"In its extreme form, a centralized system locates all processing within a single computer and all systems development work within a single organizational unit" $\boxed{1}$. The data base is similarly concentrated. Therefore, this type of processing maintains the entire data base at a single central location. Also, in a completely centralized system, the staff not only provides all technical services; it also plays the most important role in setting development priorities, standards, and allocation existing computing resources among competing demands.

4.2 Decentralized Systems

The opposite extreme from the completely centralized system is the completely decentralized system in which no communication links or coordination exist among multiple stand-alone computers. Each organizational unit having a computer supports the system with its own resources and is totally independent and responsible for all development and operation.

4.3   Hybrid Systems Proposed for This Project

Systems rarely perform in the pure centralized or decentralized
fashion; usually they have some elements of both. The approach presented
here is a hybrid system, with performance between centralized and decentra-
lized systems. Also, this approach is based on the current system features
of the bank, as well as with the witnessed trend which indicates the
possible future information needs. Likewise, this approach is presented
according with the different types of distributed systems $\lceil 2 \rfloor$.

4.4   Alternatives for Distribution of Processing

Among the different alternatives for distributing an information
system, three system types were taken into consideration for the final
system proposal; these alternatives are the following:

1) Distributed processing without local data base

2) Distributed processing, partitioned data, homogeneous software,
   heterogeneous hardware

3) Distributed processing, replicated data, homogeneous software,
   heterogeneous hardware

All three alternatives are described in the following sections.

4.4.1 Distributed Processing Without Local Data Base

This type of processing is the most basic form of distributed pro-
cessing systems. In this system some local processing takes place but
without support of a local data base; any auxiliary storage (e.g. cassette
or floppy disk) is used to store some programs and transactions. Edited
transactions are usually stored temporarily in a transaction file, and
transmitted periodically (e.g. daily) to the central computer for further
processing (see fig. IV.1).

(19)

CENTRAL DATA BASE WITH DISTRIBUTED PROCESSORS
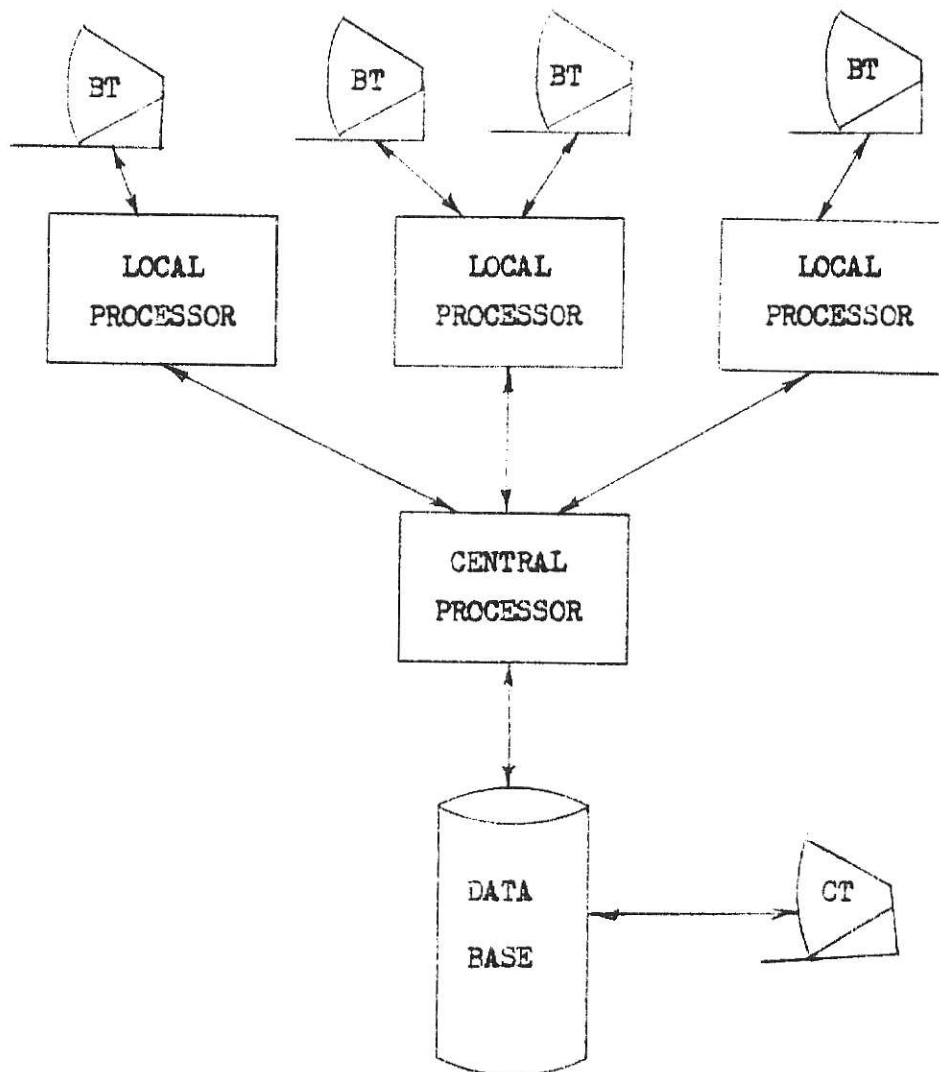
CT = Central Terminal
BT = Branch Terminal

(Fig. IV.1)

In this project, this alternative represents not only the cheapest alternative, but also the most feasible approach for implementation. This is due to experience with the performance of a current (and well known) processor, we can make a good hardware system selection for each regional office. Along with this we also can design the software required based upon a mature software design.

This alternative will permit some decentralized processing. Data capture, data transmission and consulting processes will be done through regional processors. Now each regional office not only is going to collect source documents and queries, but also is going to be responsible for the data processing carried out in it, as well as the quality of data that it is going to transmit daily to the central site. The data preparation and data base update processes will remain as in the current system, because local data bases are not allowed in this type of processing. Instead a central data base is going to be maintained.

If we compare this alternative against the current system we may observe that problems such as loss of information (because source data travels by bus or plane long distances in order to be processed), late delivery of source data (because transportation problems), correction of errors, and bottleneck presented at data capture and consulting processes are going to be eliminated.

Control of source data and opportunity of information also are going to be improved not only because each office will be responsible for its local source data, but also because source data is going to be trans-mitted daily to the central site instead of monthly.

4.4.2 Distributed Processing, Partitioned Data, Homogeneous Software, Heterogeneous Hardware

In this type of processing, data is divided and allocated across several computers. Data bases are usually partitioned based upon data locality and data accessibility; this means that files are allocated on the machine that most likely will receive the heaviest usage (see fig. IV.2) Therefore, this type of distributed processing permits not only local processing but also the maintenance of a local data base. The cost of the necessary hardware is almost always larger for a distributed system than for a centralized system. However, the use of mini-computers may in some cases yield a smaller cost for the distributed system. Also the use of this type of processing is attractive when there is a low probability of having to access a remote data base. In this system updating represents an easy task, because information is controlled and modified at only one location. Response time and availability of the system is increased. Usually, communications become reduced. However, in the bank's study case sometimes it is required to have information about a user who may have different loans at different locations. In this case communications have to be done through the whole network.

Finally, this type of processing represents not only a potential problem when data base administrators intend to obtain agreements about standards, but also the overall organizational development may become reduced and confused due to lack of centralized planning and coordination. [6,7]

DISTRIBUTED PROCESSING, PARTITIONED DATA,
HOMOGENEOUS SOFTWARE, HETEROGENEOUS HARDWARE

CBT = Central Branch Terminal

RBT = Regional Branch Terminal

(Fig. IV.2)

(23)
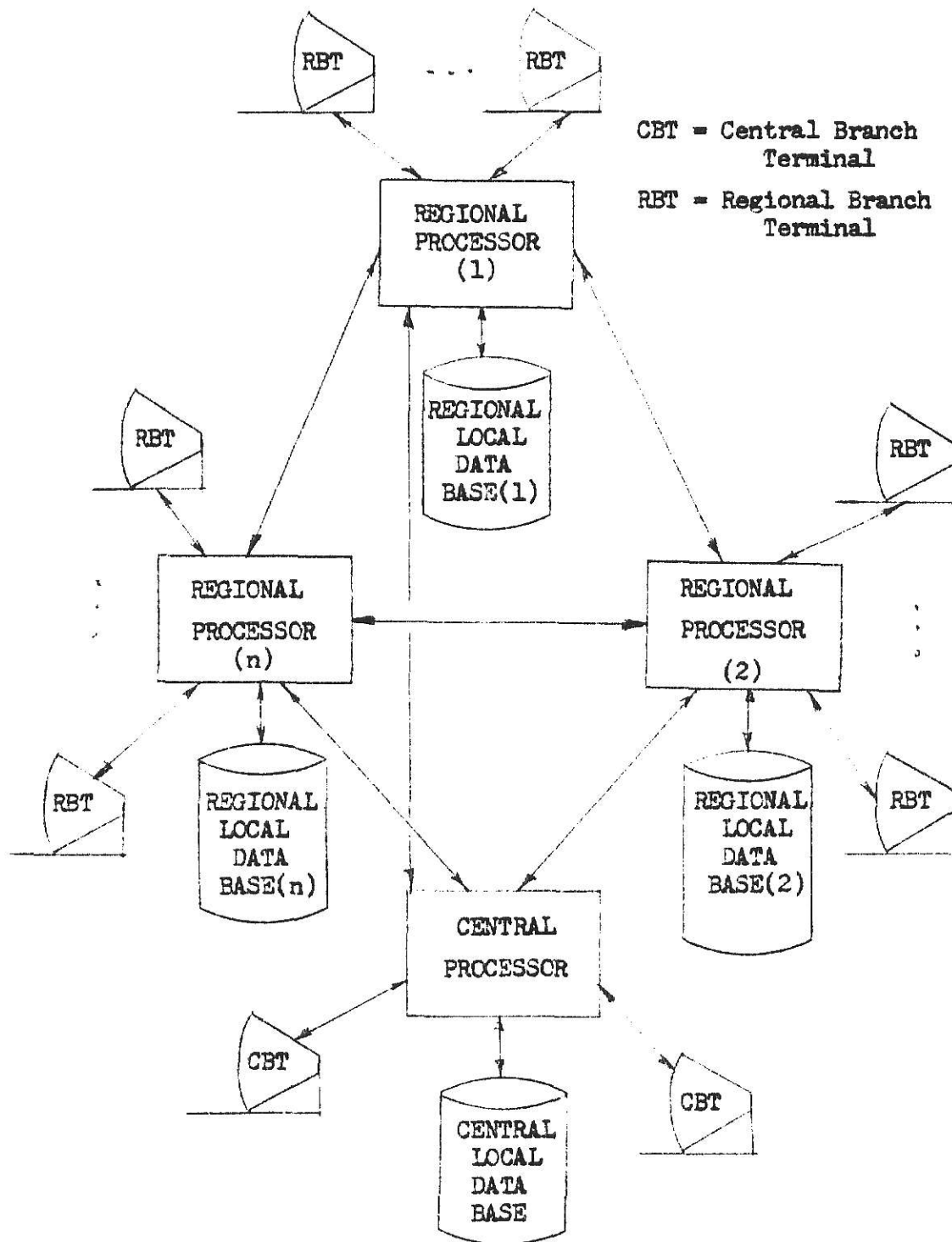
4.4.3 Distributed Processing, Replicated Data, Homogeneous Software, Heterogeneous Hardware

The overall purpose of this type of distributed processing is to divide the data and allocate the resultant portions of data across several computers. Data bases are usually partitioned based upon data locality and data accessibility; this means that files are allocated on the machine that most likely will receive the heaviest usage. However, some data must be allowed to be replicated in two or more computers.

Taking into account the geographical distribution of bank's branches, the data locality and data accessibility presented in the system, as well as the required centralized development; the distributed system that best fits into the bank's characteristics, is that which allows local data processing in order to maintain a local data base, and also provides summary data which are periodically transmitted to the central site; that handles system wide processing and maintains a central data base as well as overall systems development (see fig. IV.3).

4.5 Description of the Operations in Distributed Systems

In order to be consistent with the description of the current system, this section has been structured into three distinct modules: data capture, data preparation, and data base consultation.

4.5.1 Data Capture Description

The following description is going to be applied to the three different alternatives, not only because data capture procedure is going to be common among the three alternatives, but also because this stage represents the first step in the development of a distributed system.

DISTRIBUTED PROCESSING, REPLICATED DATA,
HOMOGENEOUS SOFTWARE, HETEROGENEOUS HARDWARE
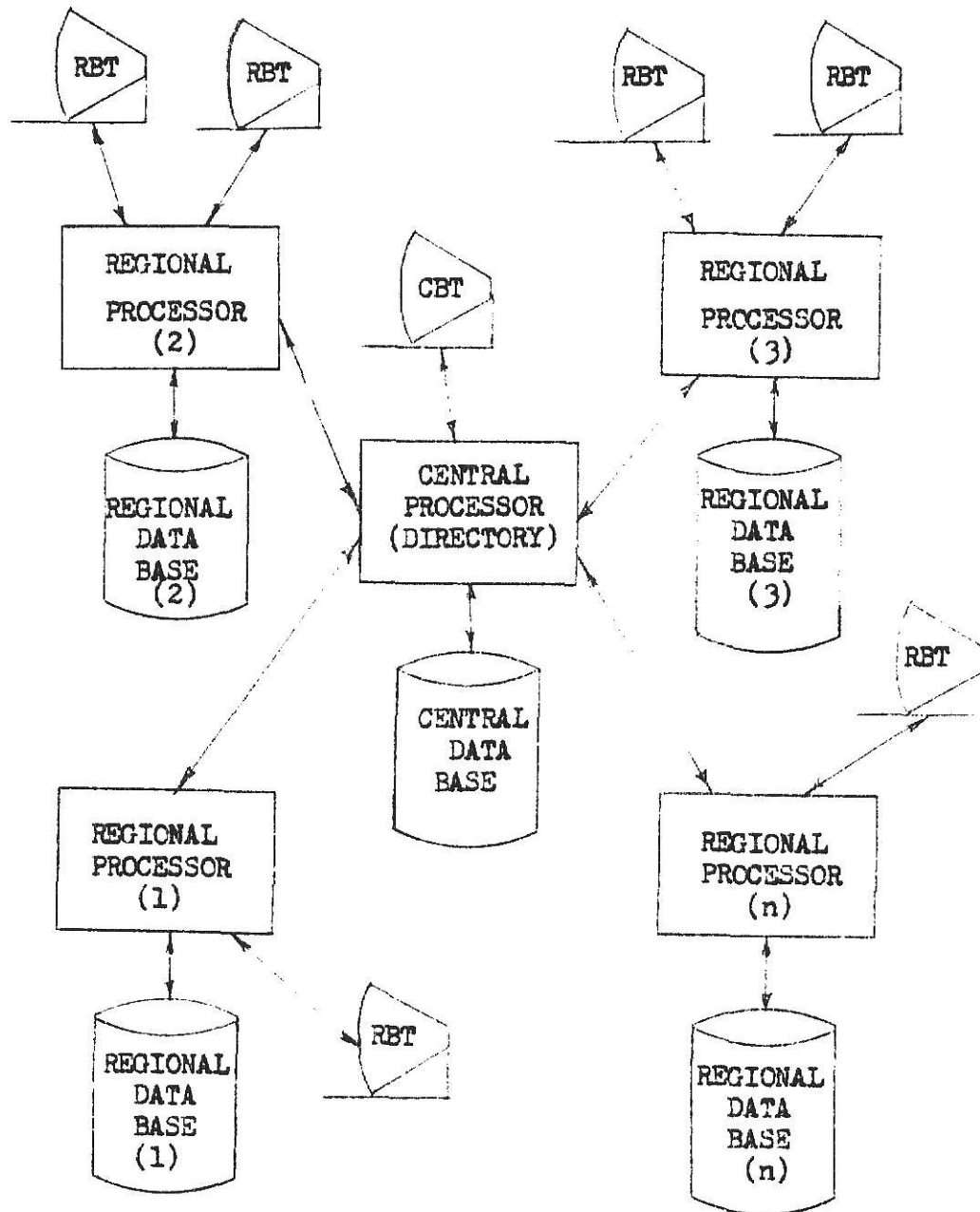
(Fig. IV.3)

The activities concerned with data capture are going to be quite different compared with those in the current system; because most of the clerical and source documents shipping operations will be eliminated. The only operations that remain the same are the procedures that are followed in order to grant a loan (such procedures are outside the scope of this report), and the coding of the source data.

Now at each branch office, instead of keeping coded forms for up to one month, each branch has to transmit the information every day to its corresponding processor (see fig. IV.4).

Transmission of source data from branch office to its corresponding regional or central office might be performed by a keyboard-terminal operator based on established format and the use of a terminal located at each branch office.

4.5.2 Data Preparation Description

As it was shown, in the current system all data preparation activities are carried out at the central office. For the distributed system, data preparation is going to be distributed across regional offices and the central office (see fig. IV.5).

If we contrast the current system with the proposed system (see figs. III.3 and IV.5) it is easy to realize a substantial reduction of processes, and an enhancement of information flow. This is because most of the clerical operations and source documents shipping operations will be eliminated.

Data preparation stage is going to be constituted by the following processes (see fig. IV.5):

(Fig. IV.4)

(27)

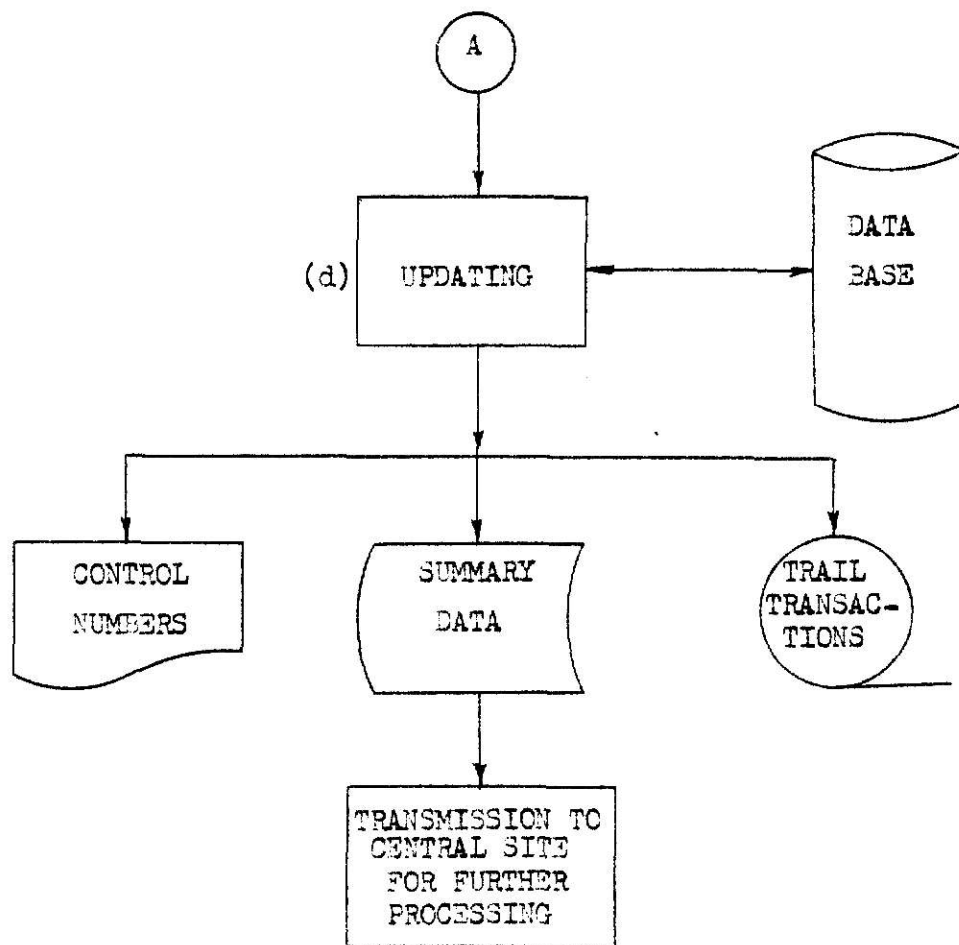# DATA PREPARATION

BT

BT

BT

RECEPTION
(R.P.)

(a)

R.P. = Regional
Processor

DATA OK

CHECKING FOR
NEW USERS

(b)

OLD USERS

NEW USERS

TRANSMISSION
TO CENTRAL
COMPUTER

(c)

ASSIGNING
KEY SYSTEM

NEW USERS
WITH KEYS
SYSTEM

TRANSMISSION TO
REGIONAL
COMPUTER

A

(Fig. IV.5)

(28)
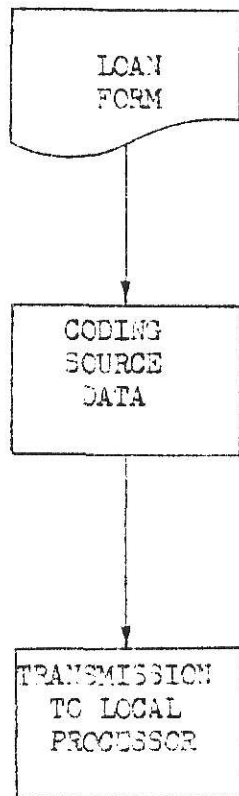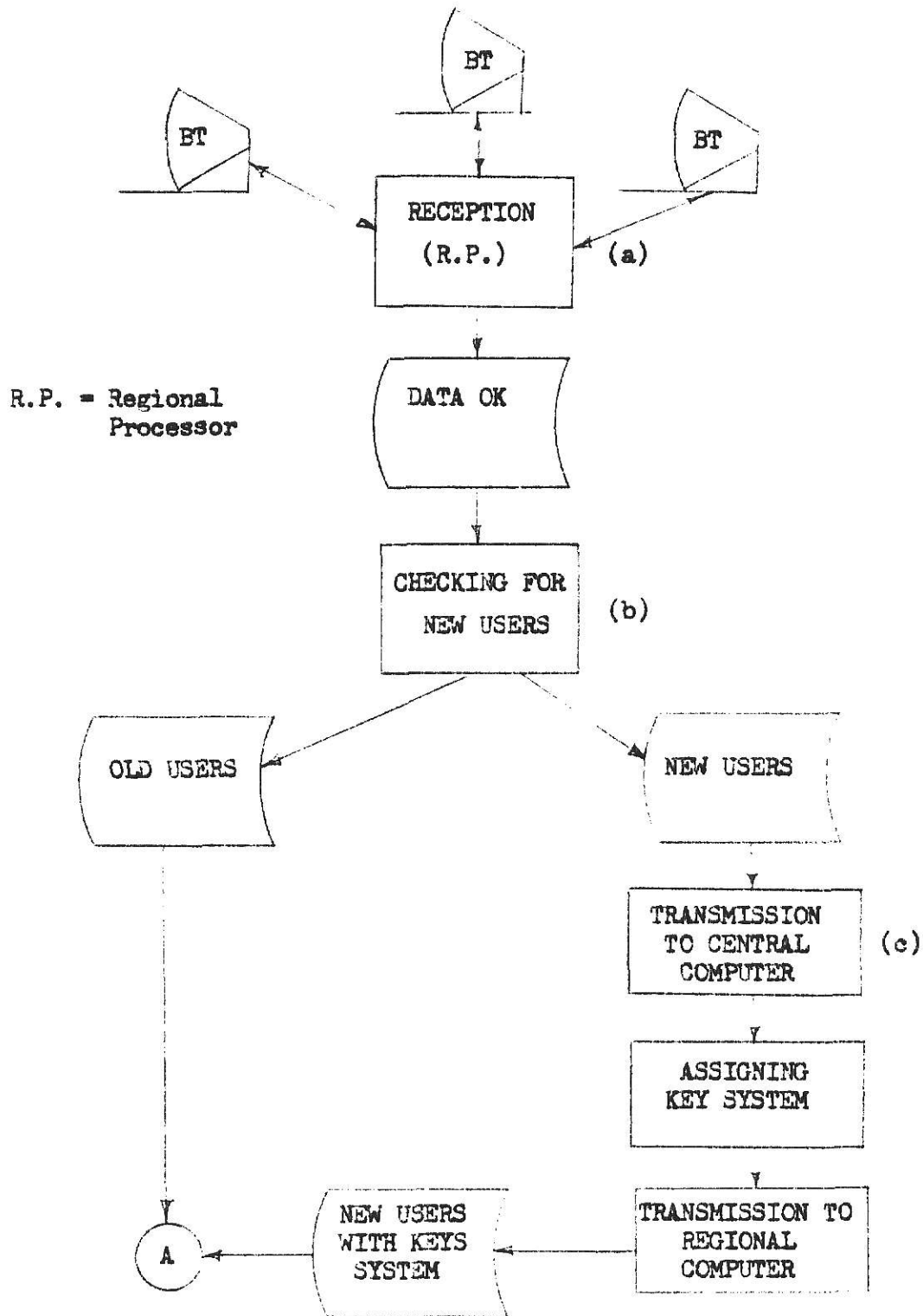
BT = Branch Terminal

R.P = Regional Processor

(Fig. IV.5) **continued**

a) Reception. Only for alternative 1, the daily transactions file is going to be transmitted entirely to the central processor for further centralized processing. This process is going to be the same for the other two alternatives. Therefore the following description is going to be applied to all of them.

Each processor will multiplex the reception of source data that is going to be sent from terminals. Software at each local processor will handle such reception of information. Also verification of source data is going to be performed in the same process by the same person who has typed source data on the terminal. Verification process is going to be carried out right after source data has been received by the processor and allocated in a work area. Software will make possible the displaying back on the same terminal the information just transmitted from that terminal, and also software will request to enter the same source data again in order to verify this information with that contained in the work area. The result of this process is a disk file with daily transactions ready to be checked for new users in the system.

b) Checking for new users. From this point until the end of this stage, alternative 1 will continue making use of the remaining and existing processes in the current system (see fig. III.3), because this alternative does not allow the continuation of distributed processing in order to maintain a data base.

On the other hand, the remaining data preparation procedures are the same for alternatives 2 and 3. Thus, the following description will be applied to both of them.

Once the source data have been checked and are acceptable, the

obtained file is going to be searched in order to generate two files. One file is going to contain all those users that are already in the system, and another file is going to contain all new users who have just received their loan from the branch. This second file is transmitted to central site.

c) Transmission from regional processor to central processor. This process is optional, and it is executed when the sender is not the central processor. This is because every user has to have a unique item in his record, which is a unique key assigned and controlled at the central site. This policy of centralization is needed in order to avoid duplicity of keys.

If the heterogeneous hardware of the regional processor produces code not compatible with the ASCII code used in the central computer, then before sending or receiving data to or from the central computer, some data translation has to be performed in order to provide compatibility between processors. This data translation is concerned only with word length and character representation. The transmission from regional processors to central processor must be initiated according to a scheduling policy. (This scheduling policy indicates that there is no need to transmit at any time).

It is important to point out that data transmission from regional processors to central processor, involves only that information that is necessary to be at the central site in order to proceed with the assignment of unique key. It means that the original length of records held at every regional processor are going to be reduced before sending to the central site in order to avoid unnecessary data transmission.

Once all the information corresponding to the new users in the system has been received, the assignment of unique keys is going to be executed in batch mode at the central site. As a result of this process a sorted file is going to be obtained (by regional processor number), containing the information input plus the unique keys (see fig. IV.6). After this file is generated, the central processor initiates the transmission back to every regional processor (Only the processors responsible for the particular data).

d) Updating. When each regional processor has received all information back, then the updating process is executed at each processor in batch mode. Updating the data base is performed using as input files the new users file (just received) and the old users file obtained at process b. The actions that are going to be taken for updating are the following:

1. If a record corresponds to a new user in the system, then this record is added to the data base.

2. If a record corresponds to an old user and if the record also corresponds to the last payment, then that record is deleted from the central data base, but is kept in the local data base for further studies.

3. If a record corresponds to an old user and if the record does not correspond to the last payment, then the existing record in the data base has to be modified with the amount indicated in the input record obtaining the new balance for that user.

As a result of updating process the following is obtained:

a) Updated data base ready to be consulted.

b) A listing of the numbers of new records in the data base, of deleted records from the data base, and of current records that are contained in the data base.

RECORD FORMATS

FIELD NAME	BRANCH NO.	CLIENT NO.	CLIENT NAME	DATE			LOAN AMOUNT	PAY-MENT TYPE	ECO-NOMIC AREA	SPECIFIC ECONOMIC SUBAREA	LOAN APPLIED AT			
				M	D	Y					State No.	County No.	City No.	ZIP
TYPE	INTEGER	INTEGER	CHARACTER	INTEGER			INTEGER	INT	INT	INT	INT	INT	INT	INT
LENGTH FIELD	4	4	30	6			8	1	1	4	2	3	3	5
POSITIONS	1 - 4	5 - 8	9 - 38	39 - 44			45 - 53	54	55	56 - 59	60-61	62-64	65-67	68-72

ORIGINAL RECORD

BRANCH NO.	CLIENT NO.	CLIENT NAME

RECORD FOR ASSIGNING SYSTEM KEY

BRANCH NO.	CLIENT NO.	CLIENT NAME	SYSTEM KEY

RECORD WITH SYSTEM KEY ASSIGNED

(Fig. IV.6)

(33)

c) A summary file that will be transmitted to the central processor for further processing so as to maintain a current data base at the central site. This summary file is composed of records corresponding to new users in the system, also records that indicate deletion of records, and finally one record for each customer that has more than one loan (presenting a global amount).

d) An audit trail tape containing all performed transactions as well as check points. This tape will be used for system recovery from possible software or hardware failure.

e) A file containing all deletions. This file will be used for historical and statistical purposes.

4.5.3 Consulting Data Bases

For alternative 1, since there is going to be only one central data base, this process is going to be performed accessing the central data base directly from each regional branch terminal, based on established formats and according to a scheduling policy.

Since data locality and data accessibility are characteristics of the information system, in alternatives 2 and 3 most of the consulting process will be performed on local data bases. However, sometimes it is required to know information about a particular user in the whole system. In order to accomplish this requirement, in alternative 2 it will be necessary to have communication lines among all processors so as to permit a complete search through all local data bases (see fig. IV.2). In alternative 3, since the central site maintains a central data base, the same requirement is accomplished by accessing directly the central data base from the demanding terminal. In this alternative it will not be necessary to have communication lines among all processors; instead communication lines will be only connecting the central processor with each regional processor (see fig. IV.3).

(34)

Also it will be possible to access a regional data base from a central site to provide detailed information about a remote user at the central site (see fig. IV.7). In order to be able to access a particular regional data base from the central site, it is necessary to have a data directory at the central site, so as to permit an easy look up of the data base containing the required information.

Finally, in order to provide security to the system, it is advisable to make a study for authentication of users, in order to allow access to data bases by only those users that have passed the authentication process.

CONSULTING DATA BASES
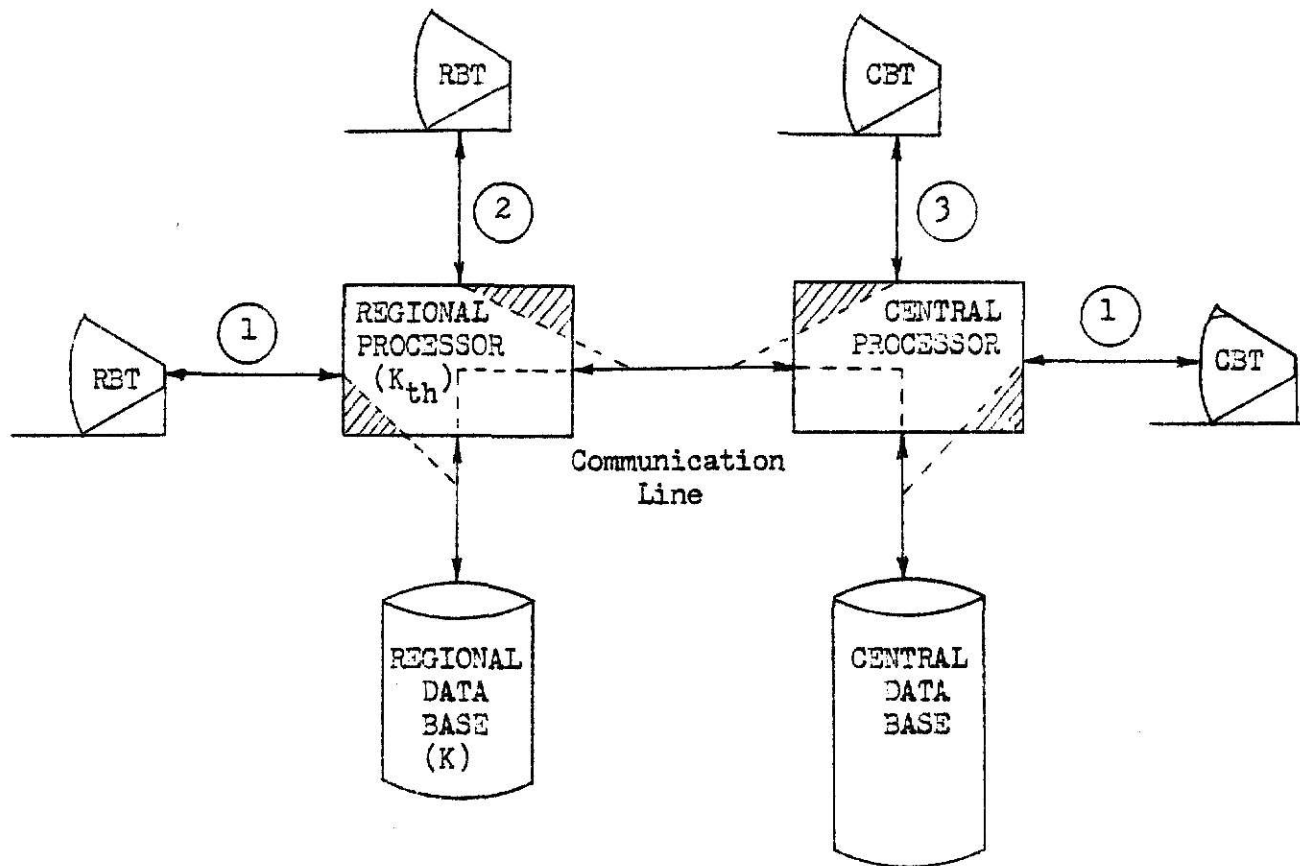
(1) Query to local data base

(2) Query to central data base

(3) Query from CBT to Regional Data Base

RBT = Regional Branch Terminal

CBT = Central Branch Terminal

Based on user desire, hardware switch decides which data base is going to be accessed.

(Fig. IV.7)

(36)

4.6 Additional System Requirements.

One of the most important factors that has contributed to the bank's management decision for decentralization, is that according to the current and future information needs of the bank, and to the current state of the art, a fantastic rate of development of computer and communications technology are opening a vast new opportunity for implementing more cost-effective information systems. New system components are going to be required in order to be able to implement the new type of data processing system.

a) For the three alternatives it will be necessary to have at each regional office a mini-computer. Therefore, it will be required to acquire one for every regional office that does not have one already. Due to the objective of this master report, the size and system configuration for these hardware systems have to be studied in a separate project. However, it is going to be mentioned that the characteristics of the processor have to include among others, to handle direct access devices for secondary storage, support high level languages such as Cobol or PL/1, support time sharing, and multiprogramming.

b) Also for the three alternatives each office has to be supplied with at least one non-intelligent terminal (e.g. TTY, CRT, etc.) in order to be able to transmit or to receive to or from their corresponding processor.

c) For alternatives 2 and 3, only one type of DMS will be used for the network information system. This is because the simplest form of a distributed data base occurs when all the processors in the network use the same DMS. No translation from one language to another is then needed, but some minor inconveniences are caused by hardware incompatibilities

such as different word length and character representation between computers. Special software or hardware is needed for communication between the DMS at the different processors in the network. Ideally, the information stored in the network can be viewed by the user as a large data base.

d) In the three alternatives it will be necessary a communications system in order to permit the flow of information among central computer and terminals.

e) Personnel have to be educated or have to be hired in order to implement and maintain the new data processing system.

f) For alternative 3, it is very important to point out that it could be wise to have two computers at the central site; not only because bottlenecks are avoided at the central computer, but also because in this way we provide reliability to the system. In this case, one computer can be processing summary data coming from the local computers in order to maintain a central data base, and the other computer may be processing the local information generated in the central area so as to maintain a local data base. Reliability is obtained because if the central computer fails the second computer (local computer) can be used in order to provide a continuous service to requesting regional (remote) processors.

V. Benefits of the New System

In order to justify the implementation of the distributed system, the following expected benefits are going to be explained. First of all, it is important to point out that the three different alternatives meet in some extent the fundamental requirement for decentralization.

5.1   Analysis of Alternatives.

As we may see, alternative 1 is the most basic form of distributed systems. In this project, this alternative represents not only the cheapest one, but also the most feasible for implementation. This is because taking into consideration the experience about hardware's performance of a current processor, we can make a good hardware system selection for each regional office. Along with this we can also design the software required based upon a mature software design. Nevertheless, according to the future information needs of the bank, this type of processing may become obsolete soon. Therefore, this alternative can be viewed as a possible initial step in the development of a more complex alternative. Meanwhile, management and staff become more confident and mature respectively in the use of this new type of processing.

In alternatives 2 and 3, data is divided and allocated across several computers based on data locality and data accessibility presented in the system. These alternatives permit not only local processing, but also the maintenance of a local data base. The cost of the necessary hardware is almost always larger for a distributed system than for a centralized system. However, the use of mini-computers may in some cases yield a smaller cost for the distributed system.

Alternative 2 could be very attractive not only because the low probability of having to access another remote data base, but also because updating represents an easy task since information is controlled and modified at one location. However, in the bank's study case sometimes it is required to have information about a user who may have several loans in more than one region. In this case communications have to be done through the whole network. Also alternative 2 represents not only a potential problem when data base administrators intend to obtain agreements about standards, but also the overall organizational development may become reduced and confused due to the lack of centralized planning and centralized coordination.

Alternative 3 allows local data processing in order to maintain a local data base, and also provides summary data which are periodically transmitted to the central site, that handles system wide processing and maintains a central data base. Thus this alternative also will allow data replication.

In this alternative the communication required in alternative 2 in order to obtain information about a multiuser (*) becomes reduced, because at any branch it is only needed to access directly the central data base in order to obtain complete information about that user.

Taking into account the above, and since decentralization in this project means decentralization of resources and operations based on a centralized overall systems planning and systems development; alternative 3 seems to be the one that best fits into the bank's characteristics.

(*) Multiuser means a user who has more than one loan, and each loan could be granted at different regions.

The current problems presented in the bank, such as loss of information (because source data travels by bus or plane long distances in order to be processed), late delivery of source data (because of transportation problems), correction of errors, and bottleneck presented at data capture, and consulting stages are going to be eliminated.

Control of source data is going to be enhanced not only because this activity is going to be shared among offices, and each office will be responsible for its own local information, but also data volume will be easily handled because source data is going to be processed daily.

Availability of information is another system enhancement. If we contrast the current system with the proposed one, we may see that updating process in the current system is performed every month and in the distributed system updating process will be executed daily at each office. Therefore, if a user obtains a loan and attempts to obtain more than one loan during the same month, then this situation will be detected.

By the same token, if we continue contrasting systems, then we will see that the expected response time will be tremendously reduced because communications become reduced since the required data usually will be allocated at the local site.

Another benefit obtained from the distributed system is that if hardware or software systems fail at one processor, this fact will not cause failure through the whole system. Therefore, machine and site independence will improve the integrity of the system.

It is important to realize that the entire system is going to be well supported. Because if a hardware system fails, then that system will have near to it as back perhaps more than one hardware system equal

or similar. Software systems can be supported by people that may be moved from one office to another in order to substitute for ill or promoted personnel.

Decentralization in this project is understood as decentralization of resources and operations based on centralized overall systems development. Then, for alternatives 2 and 3, maintenance of local data bases and system operations are going to be performed accordig to standards that are going to be issued by staff at the central office. Thus, the potential problem presented by the possible power that a data base administrator may acquire in the organization, become reduced or eliminated.

Along with these benefits some tradeoffs have to be done, in order to meet the expected system performance. For instance in alternative 3, replication of data is going to be needed so as to provide rapid backup in case of failure of a device, decrease communications and/or dependencies on the communication facilities between information processors. However, updating becomes complex because a distributed system in which there is duplication of data between different data bases raises the problem of maintaining consistency of data at the different data bases particularly after a system failure.

Finally, since the purpose of this master's report is to present insights, requirements, and conditions that may be considered in the implementation of a distributed information system; it must be mentioned that three studies have to be done in order to provide management with information necessary to plan system implementation. One study is the costs study that will be in function of the second study, which will concern the design of hardware and software systems that are going to be implemented at each location in the network. The third study has to be

(42)

done in order to obtain the new organizational structure as well as the new job descriptions for the new structure configuration.

VI. Conclusion

One of the main problems facing large organizations today concerns increasing their ability to deal with operational complexity more effectively. The fantastic rate of development of computer and communications technology opens vast new opportunities for implementing distributed systems in order to alleviate and reduce operational complexity.

A new advance in technology can be applied badly just as readily as it can be applied well. This is especially true of distributed processing systems because of its lack of maturity.

This master's report is intended to present through a sample system insights, requirements, and conditions that should be taken into consideration when a bank is going to initiate a conversion from centralized processing to decentralized processing.

Successful implementation of distributed systems requires the development of increased maturity about software and hardware technology, and systems management to develop managerial skills to comprehend and cooperate in the achievement of such successful implementation. **A cost/benefit analysis should be accomplished as a next step.**

VII. Bibliography

1. Aschim, Frode. "Data Base Networks-An Overview". Management Informatics (3,1), Feb. 1974, pp. 12-28.

2. Booth, G.M. "Distributed Data Bases: Their Structure and Use". Info-Tech State of the Art Report, Distributed Systems, 1976, pp. 201-213.

3. Casey, R.G. "Allocation of Copies of a File in an Information Network". Proceedings AFIPS SJCC, volume 40, 1972, pp. 618-623.

4. Chu, W.W. "Optional File Allocation in Computer Networks". Computer Communication Networks, (Kuo, F.F., Editor), Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.

5. Chu, W.W. "Performance of File Directory Systems for Data Bases in Star and Distributed Networks" Proceedings AFIPS NCC, volume 45, June 1976, pp. 585-593.

6. Comba, P. "Needed Distributed Control". Proceedings of the International Conference on Very Large Data Bases, volume 1, No. 1, Sept., 1975, pp. 364, 372.

7. Down, P.J. and Taylor, F.E. "Why Distributed Computing", NCC Publications, 1978.

8. Emery, James C. "Managerial and Economic Issues in Distributed Computing". Information Processing 77, (B. Gilchrist, Editor), North-Holland Publishing Company (1977), pp. 945-955.

9. Fry, J.P., and Sibley, E.H., "Evolution of Data Base Management Systems" Computing Surveys, Volume 8, Number 1, March 1976, pp. 7-42.

10. Prywes, N.S. "Structure and Organization of Very Large Data Bases", F. Gruenberger, Prentice-Hall, Inc., Englewood Cliffs, N.J. 1969.

11. Slonim, J., Schmidt, D., and Fisher, P., "Considerations for Determining the Degree of Centralization or Decentralization in the Computing Environment", Information Management. Volume 2, Number 1, February, 1979, pp. 15, 30.

APPENDIX A

CODING LOAN FORM

BRANCH NO. __ __ __ __   BRANCH NAME _____

CLIENT NO. __ __ __ __   CLIENT NAME _____

DATE   M M D D Y Y   AMOUNT _____
       __ __ __

PAYMENT TYPE   △   △   △
       EVERY   Month   2 Months   3 Months

LOAN WILL BE USED IN THE FOLLOWING ECONOMIC AREA:

Industrial   Agriculture   Cattle   Fishery
Production
   ◯   ◯   ◯   ◯

Buildings and   Tourism
Houses Construction   Promotion
   ◯   ◯

SPECIFIC ECONOMIC SUBAREA NO. __ __ __ __

LOAN WILL BE APPLIED AT:
  GEOGRAPHICAL LOCATION

              STATE NO. __ __   STATE NAME _____

              COUNTY NO. __ __ __ COUNTY NAME _____

              CITY NAME _____ ZIP CODE __ __ __ __ __

(46)

ANALYSIS OF A CONVERSION TO
DISTRIBUTED DATA PROCESSING

by

JOSE GENARO SERGIO SILVA LOPEZ

B.S., Escuela Superior de Ingeniería Mecánica y Eléctrica (México), 1970

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1979

Abstract

A bank is a nationwide service organization, consisting of a central office and regional offices. Regional offices are distributed across the country according to geographical locations. Each regional office controls branches that are allocated in that region.

The bank presently runs its applications at central office on a large computer. Over the years both the bank's business and organization have expanded to the point where the current centralized system no longer has the ability to satisfy information needs within the organization; instead operations are becoming increasingly difficult to handle. Along with this, management has issued a general policy, indicating a strong desire for decentralization.

This master's report is intended to present insights, requirements and conditions that may be taken into consideration when moving from centralized data processing to distributed data processing.