

A  BOOK

TP274
M3

8360237

STRATEGY FOR DISTRIBUTED DATA PROCESSING

JAMES MARTIN



PRENTICE-HALL, INC., Englewood Cliffs, New Jersey 07632

684

Library of Congress Cataloging in Publication Data

Martin, James, (date)

Design and strategy for distributed data processing.

"A James Martin book."

Includes index.

1. Electronic data processing—Distributed processing.

I. Title.

QA76.9.D5M386

001.64

81-5917

ISBN 0-13-201657-5

AACR2

Design and Strategy
for Distributed Data Processing
James Martin

© 1981 by James Martin

All rights reserved. No part of this book
may be reproduced in any form or by any means
without permission in writing from the publisher.

Editorial/Production Supervision: *Lynn S. Frankel*

Cover Design: *Mark A. Binn*

Jacket Photograph: *Carl Fischer (Courtesy of
Western Electric Company)*

Manufacturing Buyer: *Gordon Osbourne*

Printed in the United States of America

10 9 8 7 6 5 4 3 2

PRENTICE-HALL INTERNATIONAL, INC., *London*

PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*

PRENTICE-HALL OF CANADA, LTD., *Toronto*

PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*

PRENTICE-HALL OF JAPAN, INC., *Tokyo*

PRENTICE-HALL OF SOUTHEAST ASIA PTE. LTD., *Singapore*

WHITEHALL BOOKS LIMITED, *Wellington, New Zealand*

**DESIGN AND STRATEGY
FOR DISTRIBUTED
DATA PROCESSING**



A James Martin BOOK

TOP MANAGEMENT
COMPREHENSION
AND COMMITMENT

APPOINTMENT
OF A CORPORATE
DATA STRATEGIST

CORPORATE STRATEGY FOR DATA BASES

Establishment of organizationwide
data base standards.
Adoption of organizationwide
data description language.
Selection of data base management
software
Data dictionary selection and
establishment.
Establishment of design procedures.

CORPORATE STRATEGY FOR DISTRIBUTED PROCESSING

Establishment of an organizationwide
standard for DDP.
Establishment of areas of control
for DDP.
Determination of how data should be
distributed and interlinked
Determination of end user
methodologies and strategy.

SELECTION OF A
DATA BASE
ADMINISTRATOR

SELECTION OF
DATA BASE INQUIRY
FACILITIES.
QUERY LANGUAGES,
REPORT GENERATORS,
DIALOGUE GENERATORS

END USER
PROGRAMMING
LANGUAGES

END USER TRAINING

END
USER
QUERIES

TOP DOWN PLANNING

Determination of what subject data
bases are needed.
Identification of data structures
including files.
Determination of geographical
location and data structures.
Mapping data structures against
application and processes.

THOROUGH TRAINING
IN DATA ANALYSIS,
CANONICAL SYNTHESIS
& LOGICAL DESIGN

ESTABLISHMENT OF A
CONVERSION STRATEGY
TO ALLOW NON-DATA-BASE
SYSTEMS TO COEXIST
UNCHANGED WITH THE
NEW DATA BASE SYSTEMS

ON-LINE
DATA
ENTRY

INITIATION OF
DATA BASE USAGE

SELECTION OF
A SEED PROJECT

PLANNED EDUCATION
AND EXPERIENCE
DISSEMINATION FROM
THE SEED PROJECT

A PLANNED SEQUENCE OF
PROJECTS FOR DATA
BASE EVOLUTION

ACCURACY
CONTROLS

DATA ENTRY BY
END USER DEPARTMENTS
DATA CAPTURE AT
SOURCE

SECURITY AND
PRIVACY CONTROLS

BOTTOM-UP DESIGN

Data analysis.	Design of
Dictionary Building.	a stable
Canonical Analysis.	logical
Schema Representation.	model of
Usage path analysis.	each subject
Physical design.	data base.

DEVELOPMENT OF
PRODUCTION
SYSTEMS

CREATION OF
SUBJECT DATA BASES

PHYSICAL DESIGN FOR
OPTIMUM PERFORMANCE

INFORMATION
QUALITY CONTROLS

8360237

THE PLANNING AND EVOLUTION OF CORPORATE DATA BASE FACILITIES.

CORPORATE STRATEGY FOR NETWORKS

Establishment of an organizationwide strategy for networking.
Determination of networking architecture(s).
Design of a stage-by-stage network development plan.

DESIRABLE COMPUTER INDUSTRY DEVELOPMENTS

MORE FLEXIBLE DATA BASE MANAGEMENT SYSTEMS (RELATIONAL?)

FULLY FLEXIBLE NETWORK FACILITIES

STANDARDS FOR DATA DESCRIPTION

BETTER DATA BASE PERFORMANCE

LAYERED ARCHITECTURES FOR DISTRIBUTED PROCESSING

IMPROVED DATA SEARCHING CAPABILITIES

DEVELOPMENT OF A SEARCH ENGINE

DATA BASE MANAGEMENT SYSTEMS IN HARDWARE

INDUSTRY STANDARDS FOR COMPUTER NETWORKS AND DDP INTERFACES

NETWORK DESIGN AND IMPLEMENTATION

EXAMINATION OF DISTRIBUTED FILE & DISTRIBUTED INTELLIGENCE REQUIREMENTS

SELECTION OF DISTRIBUTED DATA BASE ADMINISTRATOR

HIGHER LEVEL DATA BASE LANGUAGES

DISTRIBUTED DATA BASE MANAGEMENT

SELECTION OF TOOLS FOR DESIGN OF DISTRIBUTED DATA

AIDS TO DATA BASE MONITORING AND TRAINING

SELF-OPTIMIZING DATA BASE ORGANIZATION

DISTRIBUTED END USER FACILITIES

AUTOMATED DESIGN OF DISTRIBUTED DATA BASE SYSTEMS

IMPROVED COMMON CARRIER NETWORKING FACILITIES

END USER GENERATION OF REPORTS

DIRECT END-USER APPLICATION DEVELOPMENT

DEVELOPMENT OF FUNCTIONAL INFORMATION SYSTEMS

WIDESPREAD END USER EMPLOYMENT OF EXISTING DATA BASES

SEPARATION OF INFORMATION SYSTEMS AND PRODUCTION SYSTEMS

LESSENING OF SCHEDULING PROBLEMS

DEVELOPMENT OF DISTRIBUTED FILE AND DATA BASE FACILITIES

CORPORATE INFORMATION SYSTEM NETWORK



E8360237

DESIGN AND

TO CORINTHIA

PREFACE

The key to success in any enterprise is good design and strategy. This is important for the management of distributed processing. The technology of distributed processing is complex and requires a clear understanding to prevent a formula for chaos.

Distributed processing is *not* a new buzzword designed to increase the attendance at seminars or the sale of books. It has been described as inevitable as the leaves appearing on a tree in spring, and just as useful!

In this book we shall describe a clear strategy for and illustrate the techniques of good design. The benefits of distributed processing are many; the pitfalls are great.

Many factors are contributing to distributed processing and the rate of change is dramatic. Soon it will be necessary for every DP professional including senior management to be fully aware of the technology and how it can be applied. It is essential that management develop a concise strategy which identifies the implementation, if applicable, the choices and commitment for distributed processing. The impact of distributed processing in such areas as office of the future and comparison with centralized processing are discussed.

In the last section of the book, important issues such as security and auditing of distributed processing are reviewed, with recommendations proposed to deal with these sensitive areas.

Distributed processing is here to stay and will change the way hardware and software are implemented. Readers might also be interested in the author's companion book, *Computer Networks and Distributed Processing*, which deals with networks and how to connect machines together.

James Martin

**DESIGN AND STRATEGY
FOR DISTRIBUTED
DATA PROCESSING**

A  BOOK

CONTENTS

Preface xiii

PART I

POTENTIAL

- 1 The Revolutionary Change in Data Processing 3
- 2 The Impact of Distributed Processing on Organizations 16
- 3 The Changing Role of End Users 33
- 4 The Office of the Future 38
- 5 Pitfalls in Distributed Processing 73

PART II

FORMS OF DISTRIBUTED PROCESSING

- 6 What Is Distributed Processing? 87
- 7 Function Distribution 105
- 8 Hierarchical Distributed Systems 120
- 9 Horizontal Distributed Systems 128

PART

III**STRATEGY**

- 10 Strategies for Distributed Data Processing 149
- 11 Control of Complexity 163
- 12 Problems of Incompatibility 173
- 13 Centralization vs. Decentralization 181
- 14 Cost and Benefit Analysis 204
- 15 Distribution and Politics 219
- 16 The Bureaucratic Imperative 230

PART

IV**DESIGN OF DISTRIBUTED DATA**

- 17 Distributed Data 243
- 18 The Location of Data 260
- 19 Multiple Copies of Data 272
- 20 Conflict Analysis 287
- 21 Data Base-Management 307
- 22 Distributed Data Bases 322
- 23 Distributed Application Configuration—
A Qualitative Method 347
- 24 Distributed Application Configuration—
A Quantitative Method 363
- 25 Strategy and Management for Distributed Data 375

PART

V**SOFTWARE AND NETWORK
STRATEGY**

- 26 Software and Strategy 411**
- 27 The ISO Seven Layers 419**
- 28 Architecture Interfaces 441**
- 29 Physical Link Control 461**
- 30 The Transport Subsystem 469**
- 31 The Session Services Subsystem 490**
- 32 Network Management 508**
- 33 Layer 2: Interconnection Message
and Transaction Design 525**
- 34 Future Software Requirements 533**
- 35 Strategy and Management
for Corporate Networks 559**

PART

VI**SECURITY AND AUDITABILITY**

- 36 Security and Privacy 583**
- 37 Auditability 608**
- Index 619**

PART **I** **POTENTIAL**

1 THE REVOLUTIONARY CHANGE IN DATA PROCESSING “MIRACLE CHIPS”— TIME MAGAZINE

INTRODUCTION

A revolution is taking place in data processing—again.

Soon after chess-playing machines began to appear in stores, a friend showed me one he had just bought. It was one of several such machines on the market, about the size of a large pocket calculator. It could be set to play at one of six levels of skill. At level 1 it was easy to beat; at level 3 quite difficult; at level 6 it took 24 hours to make each move, blinking its lights menacingly all the while.

Since neither of us could be considered the world's greatest chess player, we decided to tackle the machine jointly. We began after dinner, playing level 3. The machine took about 20 seconds to make each move and longer when we got it worried. In deciding on our moves we took a minute or more debating the alternatives and pointing out dangers to one another. Many hours and half a bottle of Scotch later, the sinister gadget beat us.

The point of this story is that three years previously a game like that would have required one of the world's largest computers. Logic circuitry has been dropping in cost at a phenomenal rate and will continue to do so for the next decade or more. Many programs that require the world's largest computers today will run on small computers three years from now. If there is a large enough market for them, they can be put on sale at low cost, like chess-playing machines.

This breathtaking change in technology is plunging the data-processing community into an era of revolution, in which the methods of the past are no longer necessarily the right methods. From now on all data-processing managers and analysts have to consider the extent to which they will use *distributed data processing*.

Until the mid-1970's most computing was carried out by systems employing a large, centralized computer. The *central processing unit* (CPU) had a diverse collection of relatively simple machines attached to it, some of them connected via telecommunications links. For certain applications networks were built in which

large numbers of simple terminals were connected to a central computer system. Sometimes the central computer system had two processors for reasons of reliability, but the processing of each transaction was done by one large computer.

The 1970's brought the era of minicomputers, and later microcomputers. With LSI (large-scale integration) circuits the cost of building a processor dropped steadily until it became clear that one computer system could be comprised of many processors if this were useful. At the same time there was a change in the perception of how computers should be used. The concept of an isolated, factorylike machine room processing batches of data for many users gave way to users wanting their own terminals and processing capability. In some cases the users' local processing machines were connected to a distant, larger machine which maintained a data base and provided extra processing power if needed.

By the mid-1970's requirements for a new type of system architecture had become clear. A new generation must provide a stable foundation for on-line, transaction-driven data-base applications. Unlike architecture designed for batch processing, it must be highly reliable because on-line users become very frustrated if their system has frequent periods of failure. It must meet a diverse range of processing requirements. Furthermore, as technology is changing very rapidly, the architecture must facilitate the introduction of new technology without a major system disruption.

Most important, a very high proportion of people must be able to use the machines. One major study indicated that by 1985, 70% of the U.S. working force would work with computers for at least some portion of their work activity. Most of these people would not have programming skills, nor should we expect them to be able to remember mnemonics or fixed sequences of input, as many of today's operators do.

Computing, then, must change its image. It must be regarded as a communications medium, in the same sense that television, newspapers, telephone directories, and shopping catalogs are communications media. Electronic information systems are a more powerful and flexible communications medium than any of the above, increasingly vital to industry and office work everywhere.

To meet these requirements microcomputers (like the chess-playing machine), minicomputers, and large computers will all play a part. The new and difficult question is, What mix of these will provide the most effective computer resources?

New computer systems are likely to interconnect many processors, large and small. They will often be interconnected over large distances with computer networks. The term *distributed processing* implies multiple processors, usually interconnected by telecommunications.

The main promise of microprocessors and minicomputers is that they will bring computing power to vast numbers of end users. In doing so they compete with the use of terminals connected to traditional data-processing systems. Who wants to use a terminal or time-sharing service to do a job that could be done on one's own pocket calculator? However, in most of its uses a computer does not compete with a pocket calculator. It collects and disseminates information. The question of where