Issues and methods for practical distributed data processing applications—I

by MAURICE BLACKMAN

Arthur Andersen & Co.

Houston, Texas

and

HUGH RYAN

Arthur Andersen & Co.

Chicago, Illinois

ABSTRACT

This paper presents methods for one of two key activities in the creation of practical distributed data processing (DDP) systems for business computing. The first activity, discussed in this paper, is to select a configuration of hardware and software to support the implementation of a multiapplication plan. The second, discussed in a companion paper, is to select data distribution and manipulation approaches for one application within the limits set by the results of the first activity. To establish a justification for the methods, the paper selects a definition of DDP, discusses the alternatives to DDP for reaching the objectives of the enterprise, and identifies the design issues to be solved or avoided in a practical system for a commercial establishment.
INTRODUCTION

Distributed data processing (DDP) seems to have been selected as the computing industry’s newest panacea for business computing. DDP has very attractive characteristics, accompanied by some significant technical challenges for the pioneer. This situation would be acceptable if the decision to use distributed processing were always made by computing professionals. Unfortunately, the spectacular improvements in the price/performance ratings of small computers has led to a significant number of articles in the business press that promote the use of small computers in DDP configurations. As a consequence, the initiative in using DDP is taken by the chief executive, and the MIS director is frequently faced with a directive to “implement distributed processing.”

Principal Tasks

In our experience and that of our colleagues, there are two activities crucial to a successful installation. The first is the preparation of a plan or strategy for the disposition of functions, data, and hardware and the identification of the necessary supporting software. The second is the disposition of data and the associated choice of approaches to the timing and synchronization of processing for each specific application within the constraints set by the choices of the prior plan.

The two activities fit into the context of a complete method for systems development. They draw upon data developed in earlier tasks and provide input to tasks that follow. The two tasks are discussed because they result in key design decisions that control the impact of logical problems and technical gaps that we see in the present environment.

Each of the two papers discusses one of the two activities. The first is called strategy selection and the second data design. A method for collecting, organizing, and analyzing relevant data for each is presented. The methods, intended to be responsive to the current state of the art in DDP products, are aimed at minimizing present risk for the commercial user of computing. As logical challenges are overcome and more sophisticated products are introduced, the methods will evolve.

Issues

There are two groups of issues that can make DDP implementation a formidable task: logical puzzles that spring from the very nature of allocating processing between several processors and gaps in the available technology that make some approaches more costly than others. The second group of issues will lose their significance as the gaps are filled, but the first group may persist. In any event, the MIS director needs to adopt an approach to DDP that avoids both groups in order to succeed in creating a DDP environment within constraints of cost and time.

Definition

There are many interpretations of the term distributed data processing. To place this contribution in context, we define DDP as follows: A data processing technique that provides access to computing power for end users by means of multiple processors interacting through the planned exchange of data over communication lines.

The definition is explicit, because DDP systems have many sources of variation:

- The processing capability of the devices at the remote location
- The extent to which data are distributed to terminal locations
- The discipline used to communicate
- The number of terminal locations
- The frequency of the processing at terminal and central locations
- The frequency with which the processes are synchronized
- The compatibility of the devices used for implementation

Our definition varies from others in certain key respects. It is more restrictive than some in implying storage capability and hence data residence at each node. It is less restrictive in not being limited to online or real-time interaction between processors and in not implying any dynamic allocation of tasks.

We wish to stress the definition question for two reasons. First, a clear definition helps the DP professional explain how a directive to “use DDP” has been interpreted. Second, the definition can be used as a template or profile for comparison with a proposed configuration. A conforming configuration may suffer from some of the logical problems or gaps in technology that exist. The implementer of a conforming configuration will be alert to these issues and take steps to avoid them.

STRATEGY SELECTION

User Objectives

Our earlier definition hinted at one of the key objectives of the organization examining DDP: to provide access to com-
putting power for the end user. Distributed means that some function of the data processing department is reassigned to a person closer to the mainstream of the business.

There are many means by which that function distribution may be achieved, including online terminals and decentralized computers. The many variations of DDP fill the middle ground. It is the quality of direct access, independent of the data processing department, that is the principal objective of the requester or commissioner of DDP. We therefore take the objective of strategy selection to be the configuration of a combination of hardware and software components that delivers access while optimizing an objective function addressing performance, development cost, operating cost, delivery date, and ease of use.

**General Approach**

The strategy selection activity can be decomposed into a sequence of tasks, as follows:

- Identification of applications
- Identification of distributed functions
- Analysis and selection of a configuration of hardware and data communications facilities
- Selection of standards for data distribution
- Development of a catalogue of software components
- Selection of products

Strategy selection occurs during a planning phase. The results must stand for five, ten, or even more years. The selected strategy must be responsive to requirements that are imperfectly defined and likely to change. The selection approach must therefore rely heavily on the experience, knowledge, and judgment of the selectors. The steps here presented are precise in their identification and sequence but will be subjective in their execution.

**Application Identification**

If a strategy is to be durable, the selectors must have a good appreciation of the applications to be implemented. From the point of view of the user, computing services are more convenient if they are accessible through a single work station. From the point of view of the enterprise, product acquisition is more economical if negotiated for the long term. For such reasons, this first task identifies as many potential uses for the enterprise discovered by interview and research, such as size, growth, and markets sought, help to develop a plan with a sequence and time frame that reflects the enterprise’s priorities and resources. (See Figure 1.)

**Function Distribution**

Candidate functions capable of distribution at this time are the generic functions of a data processing department rather than the specific functions of single applications. This task sets limits on the maximum distribution in order to identify needed software support in the configuration and to establish overall quantities and capacity of equipment.

Table I gives examples of generic functions in three groups: operations, development, and management.

<table>
<thead>
<tr>
<th>TABLE I—Generic data processing functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operations</td>
</tr>
<tr>
<td>Inquiry</td>
</tr>
<tr>
<td>Data entry</td>
</tr>
<tr>
<td>Validation</td>
</tr>
<tr>
<td>Correction</td>
</tr>
<tr>
<td>File analysis</td>
</tr>
<tr>
<td>Printing</td>
</tr>
<tr>
<td>Posting</td>
</tr>
<tr>
<td>Conversion</td>
</tr>
</tbody>
</table>

Decisions about the dispersion of functions are responsive to characteristics of the enterprise and are made according to simple guidelines.

**Operations**

In general, operating functions are prime candidates and are considered first. If operating functions are not dispersed, the dispersion of any other functions is very unlikely. Operations functions are dispersed when the local business unit has autonomy in its own day-to-day operations, when its operations are on different business cycles from those of its parent or headquarters, and when its operations efficiency or effectiveness can be improved by rapid access to the results of computing. The only common exception would be a lack of the necessary skills or working conditions to operate computer peripherals.

**Development**

Development functions are also good candidates, for the reasons commonly cited for any form of end user participation in development: knowledge of the business, responsiveness of the design to user needs, and enhanced acceptance of the system. In addition, if the sites considered for function dispersion are widely separated, centralized development becomes difficult as a result of communications difficulties. Dispersed development is almost a necessity if the management style of the enterprise favors local autonomy as shown by a
INFORMATION REQUIREMENTS PLANNING CHART
PROPERTY AND CASUALTY

BUSINESS PROCESSES

STRATEGIC PLANNING

- ENVIRONMENTAL FORECAST
- MARKET IDENTIFICATION
- NEW PRODUCT PLANNING
- GROWTH PROJECTIONS
- PRICING POLICIES

PRODUCT/MARKET PLANNING

- MARKET DEFINITION
- PRODUCT DEFINITION
- NEW PRODUCT RELEASE PLANS
- SALES FORECASTING
- PROMOTION PLANNING
- REGULATORY APPROVALS

OPERATIONAL PLANNING

- MARKET PENETRATION
- GROSS SELL
- PROMOTION PERFORMANCE
- COMMISSION
- AGENCY PERFORMANCE
- COMPETITIVE ANALYSIS
- PROMOTIONAL ACTIVITY

SALES/MARKETING EXPERIENCE

- MARKET PENETRATION
- GROSS SELL
- PROMOTION PERFORMANCE
- COMMISSION
- AGENCY PERFORMANCE
- COMPETITIVE ANALYSIS
- PROMOTIONAL ACTIVITY

PRODUCT EXPERIENCE

- PRODUCT PERFORMANCE ANALYSIS
- LEGAL/REGULATORY COMPLIANCE
- IN FORCE ANALYSIS
- EXCEPTION ANALYSIS

UNDERWRITING CONTROL

- CLAIM FREQUENCY ANALYSIS
- LEGAL/REGULATORY COMPLIANCE
- IN FORCE ANALYSIS
- EXCEPTION ANALYSIS

POLICYHOLDER SERVICE CONTROL

- TIME/SERVICE PERFORMANCE
- COMPLAINT ANALYSIS
- SHAREHOLDER INSURANCE ANALYSIS
- IN FORCE REPORTING
- CANCELLATION ANALYSIS
- GROSS SELL
- PROMOTIONAL ACTIVITY
- PROCESSING ACCURACY
- UNDERWRITING EXPENSE ANALYSIS

CLAIMS CONTROL

- TIME/SERVICE PERFORMANCE
- ATTORNEY CONTROL
- ACCESS CONTROL
- COMPLAINT ANALYSIS
- AVERAGE PAYMENT
- PROCESSING ACCURACY
- CLAIMS EXPENSE ANALYSIS

FUND MANAGEMENT STRATEGY

- PORTFOLIO PLANNING
- LIQUIDITY PLANNING
- INVESTMENT PLANNING

FUND MANAGEMENT PLANNING

- FUND PERFORMANCE
- RISK EVALUATION
- PORTFOLIO ANALYSIS

FUND MANAGEMENT

- FUND PERFORMANCE
- RISK EVALUATION
- PORTFOLIO ANALYSIS

COMPETITIVE DATA COLLECTION

- COMPETITIVE PRICE
- PROMOTIONAL ACTIVITY
- MARKET SHARE

TRANSACTION PROCESSING

- SOLICITATION
- SOLICITATION RESPONSE
- POTENTIAL CUSTOMER MAINTENANCE
- ADVERTISING

POLICY MAINTENANCE

- POLICY ISSUE
- ENFORCEMENTS
- RATING
- MARKET INSPECTION REPORTS
- POLICY RENEWAL
- PREMIUM ACCOUNTING
- REINSURANCE
- DIVIDEND ACCOUNTING
- COMMISSION STATEMENTS
- REFUNDS
- EDA INTEREST
- WORKLOAD REPORTING

PREMIUM MAINTENANCE

- PREMIUM RECOVERABLE ACCOUNTING
- PAYMENT FEES ACCOUNTING
- BILLING STATEMENTS
- DIVIDEND ACCOUNTING
- COMMISSION STATEMENTS
- REFUNDS
- EDA INTEREST
- WORKLOAD REPORTING

CLAIM MAINTENANCE

- CLAIM ACCOUNTING
- CLAIM PAYMENTS
- CLAIM EXPENSE PAYMENTS
- SPLIT
- SUBROGATION
- REINSURANCE
- LITIGATION
- WORKLOAD REPORTING

PORTFOLIO/INVESTMENTS

- PORTFOLIO ACCOUNT
- INCOME ACCOUNTING
- REVENUE RECEIVABLE
- MATURED SELLING
- FUND MAINTENANCE
- FUND ACCOUNTING

Figure 1—Industry system schematic—detail

From the collection of the Computer History Museum (www.computerhistory.org)
profit center policy or if the products, markets, and businesses of the units are different. As with operations, the principal impediment to dispersion will be a lack of available skills. It should be noted that the dispersion choices being made here do not imply that all applications will be equally dispersed. The objective is to determine the degree of dispersion under the most favorable conditions so as to uncover software and equipment needs.

Management

Management functions are least likely to be dispersed, because the resources being managed are those shared by all users. Their proper selection and administration requires that the needs of all groups be blended and that facilities capable of working together be obtained. Conditions for the dispersion of management functions are the dispersion and local control of most of the assets of the enterprise and the dispersion of other service functions such as accounting, personnel administration, and research and development. Such an enterprise is really several distinct enterprises needing only occasional contact.

Hardware and Data Communications

Function distribution, although subjective and judgemental in its methods, is performed early, because its outcome may limit the hardware options considered. If only the simplest operating functions are distributed, then an online centralized solution is indicated. On the other hand, if extensive distribution of functions occurs, many strategies across the dispersed-access spectrum are possible. This task may be subdivided into the following smaller tasks:

- Selecting access styles and modes
- Identifying access locations
- Configuring peripherals at each location
- Determining the processing/communications strategy
- Configuring data storage

Access modes

Access styles and modes include choices of online or batch access techniques and such special requirements as OCR, MICR, microfiche, or badge readers. The distributed work stations' complement of devices can be deduced by examining the likely input and output media of each application.

Locations

The access locations to be given direct service must be reviewed for instances of very low volume not justifying equipment. Lack of volume may be a temporary condition. Plans for network expansion at a later time may be an important input to the more precise steps discussed later in this section.

Peripheral configuration

The configuration of peripherals at each site is the first precise step. The different types of peripherals need to be treated separately. Methods for establishing keyboard screen and printing requirements to satisfy random and scheduled demand are well established from online system experience.

Requirements for specialized peripherals must be considered carefully when unit capacity and cost are both high. Those forms of peripheral are more usually centralized for all but the largest sites.

Processing/communications strategy

We now come to the crux of the hardware question: should processing be centralized or distributed? The analysis consists of computing one-time and operating costs for alternative solutions and comparing them. The variety of alternatives—simple online terminals with a value-added network, smart terminals with a multidrop leased line network, intelligent terminals (stand-alone or clustered) with dialed WATS lines, minicomputers and mainframes with auto dial/auto answer—is usually too many. A technique of sampling, say three alternatives, followed by successive refinement, is preferred. The cost analysis must recognize that increased dispersed intelligence reduces data transmission volume and allows the use of low-tariff periods for transmission if the remote and central processing cycles can be decoupled.

The analyst must also consider geographically intermediate dispersion. There may be a case for sharing processors between several remote sites. This is an example of refining a strategy once a degree of distribution has been determined to be advantageous.

Data storage

The final step is to estimate data storage requirements if a strategy of physical distribution has been selected. This step is performed before more precise allocation of data within applications, since that cannot be done with certainty until designing each in turn. At this time we are concerned only with gross needs. Data storage capacity is needed for development as well as production, for programs as well as data, for system software as well as applications, for backup as well as mainline data, and for inefficiencies of use. In short, rather large estimates are indicated in preference to rather small.

Data Distribution Strategy

The final stage is to consider whether to place any limitations on the application designer so far as the complexity of data arrangements is concerned. Our principal target is to decide what forms of data synchronization are allowed and what forms, if any, of data directory are required.

Our first task is to determine whether master data centralization is indicated. Under current technology constraints, data needing to be up to date for all locations must be central-
ized. Other indicators include any suggesting the need for full-function database management system software. Such software needs the capacity and skills of a central site.

As a rule of thumb for strategic planning purposes, the designer should avoid data distribution configurations that require online posting within one commitment unit at more than one node. The designer should be wary even of online posting at a node other than the one at which a transaction originates. Paper II explains the background for this guideline. ¹⁰

Software Configuration

Once the decisions about function, hardware, and data distribution are made, the generic components of software must be identified.

The key components are those associated with the fact of distribution:

- Data communication, including the appropriate online or offline protocols
- Message-routing logic for bringing data into the center and putting it out again, including distribution list interpretation logic, logical address to physical line mapping, alternate routing, etc.
- Data location logic in the case of split data files, including directory maintenance
- Data conversion logic to map data characters, fields, records, files, and databases from one format to another
- Development aids in each node type expected to support local development
- Program and dictionary distribution logic if one site is developing on behalf of others
- Remote job request submission and acceptance

and many others.

This list shows how the function, data, and hardware distribution decisions have consequences in the complexity of the environment.

To complete this task with any reliability, the designer must understand the software components necessary to the functions of operating, application development, and resource monitoring. Each of these divides and subdivides until a portfolio of components is developed that may be used as a checklist.

Vendor Selection

At the end of the strategy development, the designer has sufficient information to consider the products necessary for creating the target environment. This step will have been in view throughout the development of the strategy, since limitations in products available are the reasons for some of the rules of thumb suggested.

The principal components to be obtained will be the processing systems of the different nodes and the communications subsystem between them. By far the most productive rule at this time is to obtain processors at all nodes from the same vendor so that much of the potential software complexity can be subcontracted to the vendor. Many vendors have defined and implemented network architectures of richness sufficient to support cooperative working among multiple processors. The decision to use different vendors will inevitably involve the enterprise in the definition of its own architecture and the software to implement it.

SUMMARY

The objective of this paper was to describe a sequence of steps for developing a DDP strategy. The steps have been described only to show their purpose. The material is offered to help identify the components of the strategy and the dependencies between them. The resulting strategy is a high-level plan adequate for costing, for vendor selection, and for commencing the development of any necessary software.

In Paper II, we consider the detail of data design for an application within the constraints of a strategy.

REFERENCES
