7

Data Management Concepts Important in the Reference Model

INTRODUCTION

With the establishment of plant-wide data networks, as discussed in Chapter 9, the exchange of information between departments and their databases becomes the area of principal focus in the development of the Computer Integrated Manufacturing System. The need for the integration of network, hierarchical and relational databases distributed on communications networks throughout the plant must be considered. In addition, the continuance of the realtime data acquisition function for process control and management information is paramount in a plant-wide control and information system. This chapter will discuss the types and characteristics of Database Management Systems (DBMS's) before presenting an analysis of those features as they apply to real-time decision support systems.

The global database is the area where the most work is necessary to assure the needed commonality discussed here. To date there has been little standardization of the field other than for the several de facto standards developed by the several major vendors for their own products.

The database requirements for the systems discussed here are very large - up to 20 to 40 gigabytes for the steel industry systems. Both distributed and centralized databases are required in the same systems. Distributed databases, probably at each operational computer system, will contain process plans, and other control instructions, immediate future production schedule items, current plant variable data for control applications and operator interface needs, and other immediately necessary items. The centralized database is historical in nature: production results, raw material and energy usage, quality control data, sales and shipping information, maintenance data, inventory levels, etc. Much of the historical database must be relational in nature since much of its use will be for inquiries and studies of past performance in the hope of improving future operations.

Provision must be made in each case for back-up of the data because of potential computer system outages or other occurrences affecting the validity or availability of the data and information contained therein.

DATABASES NOW IN USE IN THE PROCESS INDUSTRIES [80]

Through the years, the conventional file system mechanisms have been used, typically with FORTRAN, to store and retrieve real-time data. Sequential files are used for collection and archiving. Indexed files provide random access to the data. Their simple structure provides the speed of response needed.

Ready access to data for analysis and decisions requires data to be stored in a structure that permits retrieval by content rather than location. Data tables can be structured into hierarchical or network sets in a database for later use. The Hierarchical database permits location according to a decision tree. The Network database provides more direct access to a diverse set of users, but requires the users to know the data location beforehand for efficient access to the tables. The a priori structuring of both database types limits their efficiency in ad hoc queries for analysis and optimization.

The relational database meets the needs for ad hoc queries of data for decisions in today's dynamic competitive production environment. It permits the logical independence of the data from its physical location. It also permits the use of a high level, Structured Query Language (SQL) [42] for non-procedural, set-oriented access to the data. Users need only specify what is to be done and it processes the data as sets of elements, rather than a record at a time. The hiding or transparency of the data structure to the user imposes a processing burden on each query that must be taken into account in real-time applications.

Database Management Systems currently available cover the gamut of types from the simple file manager to the SQL-driven full relational structure. Expert Systems are being developed to assist in the development of queries. DBMS's run on platforms from the Personal Computer to the large mainframe. Large databases with transactional inputs and updates as in the financial markets require large distributed processors. The PC-based systems are primarily file managers with relational query interfaces.

REQUIREMENTS FOR REAL-TIME DATABASE SYSTEMS

DATABASE ACCESS MECHANISMS

Once the data has been collected, effective use of the information requires that it be accessed for analysis and decision-making. For those applications where the needs can be identified beforehand, direct, hierarchical or network access can be provided by linking the application to the data. Some assistance is required where the amount and type of data needed varies from time to time.

A Data Dictionary is a common method of providing rapid access. The location of the data is entered into a separate database when it is stored. The search is then limited to accessing a pointer to the data. In a distributed processing system, the data dictionary can be centralized in a single master table or remain resident in the nodes owning the data. In addition to distributing the processing load, the management and control of access can remain with the data owner.

The task of manipulating data in useful sets is greatly simplified and expedited by use of a high-level query language, such as SQL, which supports the collecting and storing of fields in multiple tables in a single operation. It also permits the user, program or person, to define the data by its properties rather than by location, which need not be known by the requestor.

A simplified form of query language which is especially user-friendly is the Query-By-Example. There the user defines the data needed by its properties and relationships in a simple form like the structured query. But, the system searches for, and fills-in the blanks entered by the user. Thus, without knowing what data is available in the database, the user can seek out and retrieve data which meets his selection criteria.

Query-by-Forms is also a simple, user-friendly form of user interface. There the sets of data for input or output, or both, are defined in the field labels on a form. The DBMS handles all the data manipulation, control and cataloging needed.

REAL-TIME OPERATING CONSIDERATIONS

Perhaps the one common attribute which can characterize the process data on the shop-floor or in the plant is that of the need for real-time acquisition and processing. Real-time data can be defined as the data which needs to be manipulated, (read, written, transferred or processed) under a strict real-time constraint. Some real-time data is said to be hard real-time data while others are called soft real-time data. Hard real-time data is characterized as that data which, if not manipulated within a certain specified time, will lead to severe consequence (e.g., a major process upset). Soft real-time data on the other hand diminishes its value greatly if manipulated after the deadline but does not cause any serious catastrophe (e.g., trends on averages of the point values).

Perhaps, the single most important characteristic of a real-time data management system is its ability to provide guaranteed response for any data manipulation request for which there is a strict deadline.

The real-time database system is the determinant of guaranteed response for data manipulation requests. The database system usually provides the key functions of buffer management, multirequest switching, transaction scheduling and monitoring, recovery and consistency management and data manipulation primitives. These functions must be provided and implemented so that they are compatible with the operations of the corresponding functions in the real-time operating system; otherwise, conflicting policies executed by the operating system and the database system will lead to violation of the real-time constraints on transaction execution.

KEY CHARACTERISTICS

In considering the needs for a DBMS for integrating the control and information systems from throughout the plant, some key characteristics are suggested.

REAL-TIME COLLECTION

Data acquisition in today's process plants is primarily performed by the process-connected control, analysis and monitoring systems. Collection of these values for use by the on-process operator and advanced control processors is an inherent part of their operation. Because they operate under real-time constraints, their data conversion, manipulation and storage capabilities are limited. Conversion to engineering units as well as time-slice or snapshot collections may need to be performed by higher-level units. The DBMS must provide an easily adaptable interface to these data sources.

TIME-TAGGED RECORDS

Time is an added dimension of the data acquired from the process. Process snapshot values as well as state value changes need to be associated with a time of occurrence for later analysis of events. The closer that the tagging of the data is to the data acquisition process, the more accurate and useful the data will be. If the process-connected data acquisition control system is not prepared to provide the time-tagging of values, the DBMS will need to support the adding of time values to data sets as well as the synchronizing of its time value with the data acquisition process. It is obvious that the determinism of the DBMS will affect the consistency of the data.

ON-LINE ARCHIVING

Once the time-tagged data have been collected, the means for its storing in coherent sets in bulk media must be provided. Here the consideration is for timely storage in large volumes compared to the structuring of the data for later access. The processing of the data into an elaborately structured database could take more time than the data acquisition cycle permits. The tradeoff is against the complexity of the query processing. Data which has a high probability of need for immediate access, such as for displays of process variable trends will need to have simple access mechanisms for timely results.

MULTIPLE USER VIEWS

The need for the control and display processing at the process-connected level will be the primary determinant of the priorities and structure of the data acquisition process. Other users in the control room or throughout the plant will need the data structured into different sets, across process units, plant areas and across time. Each of these users, whether program or person, should have his own view of the data for effective analysis and decision-making. This requires a DBMS capable of supporting those types of access to the data.

STANDARD QUERY LANGUAGE

In order to gather data in useful sets for analysis, a tool is needed to describe the data set in terms of those needs. A standard language for structuring those queries such as SQL provides those tools. Providing access to the real-time data structure is a key characteristic of the language to be used.

SIMPLIFIED ARCHIVE RETRIEVAL

Because the data collected and archived by the process-connected system is normally structured according to the variable and unit to which it is connected, as well as time, subsequent retrieval for display and analysis may necessitate complex queries and extensive processing. As with the transient user views, a DBMS capable of supporting this type of access is needed.

DATA INTEGRITY

While data collection, transmission and manipulation are in process, some of its values may not be consistent with the rest of the set. For example, if one block of the set is in error during transmission and must be retransmitted, the rest of the set is out of step with it and should not be used. Likewise, during a transaction updating a data set, other users should be prevented from using the data until the transaction is completed. Also, if the transaction should fail to proceed to completion, some means of backing-out of the data must be available. A DBMS should be able to support these data integrity needs.

Of equal importance to the integrity of the data is copying between databases. In designing for rapid local access to data, large scale copying or replication of the data should be avoided. This may compromise its integrity. An application should request no more data than it can use. To do so would imply the assumption of the responsibility for its validity, which only the owner is able to do.

ACCESS CONTROL

With multiple users able to access the database distributed across the plant, the ability of those users to affect unit operations or plant operations based on that data becomes a concern. Access to the database at each level must be capable of control by the data owner. This is the lowest level at which the data can be assured to be valid. Access to lower levels, such as the data acquisition values, must be restricted to the owner of the data, even though it may provide faster and simpler access. The ability of a DBMS to control and limit the access of particular users to certain sets or views is needed.

CHARACTERISTICS OF CURRENTLY AVAILABLE SYSTEMS

There is a wide selection of database system products on the various computing platforms that can be used at the process, plant or enterprise levels. We will briefly discuss here the characteristics of some currently available DBMS's [97, 98, 19, 25, 78] that are important in the selection of a database system at any of these levels. The implications of those characteristics for control system and plant-wide system integration will be discussed in later sections.

OPERATING SYSTEM SUPPORT

Large numbers of database systems are available on various different operating systems such as VMS, UNIX, and MVS which are typically used at the plant level. These database systems are of two kinds: one kind developed by computer system vendors and the other developed by third parties. Third party DBMS's typically run on a number of different computers, thus providing greater portability and flexibility for applications. At the process level, the database systems used are typically special purpose and supplied by control systems vendor. They typically are proprietary as are their underlying operating systems. They provide short real-time response times by integrating well with the operating system environment of the control system. Therefore, the choice at this level is quite limited currently. However, as third-party database systems support more real-time features, control system vendors will consider them for use in their systems to increase compatibility and to leverage the wide database applications already available. The choice between database systems on the main-frame operating systems at the enterprise level is smaller than that at the plant level. Database machines available as back ends to some of the host main frames provide higher performance in managing the large and complex databases present at the enterprise level.

MACHINE CAPACITY REQUIRED

The computer hardware needed to support process database management is typically based on the 68000 class of processors and 1-2 megabytes of memory and optionally supports some secondary storage devices. However, as process data grows in size, which the current trend indicates to be likely, the size of memory as well as the processor power required will increase to provide both real-time access and some ad hoc access. The machine capacity required by plant level database systems is not a serious issue because they run on so many different computer systems with different capacities. However, since many database systems at this level come as bundled packages with perhaps some unnecessary utilities and features, it is useful to consider unbundled packages which allow the users to pick and choose the DBMS related tools they need.

FILE STRUCTURE

The methods of physical and logical organization of data by current database systems cover a wide range including flat files, indexed files, relational, hierarchically structured and network structured records and files, and object-oriented structures. At the process level, real-time access is more important than ad hoc access. Therefore, organization of the data either in memory or on disk using hash pointers, indices and B-trees to improve the access to data records or fields is preferred and characterizes how many process databases are structured today.

At the plant level, ad hoc access is important. Relational databases provide the flexible database organization and access required and therefore, they are the preferred DBMS's for plant management applications. At the enterprise level, the database structures and organization methods are tied to the history of the traditional financial type applications and are more evolutionary. Use of relational DBMS's at this level will increase access to enterprise databases and promote integration of plant and enterprise data.

USER INTERFACE SUPPORT

Process databases are often accessed directly by users. They are typically accessed by control system applications as well as display systems which interface to the user. As the network interconnection of systems expands, better visibility of process data to the plant will be afforded. The sophisticated user interface tools now available will make this possible. Current DBMS products provide a variety of user interface mechanisms including SQL, Forms, Query-by-example, graphics and English like natural language query mechanisms as noted above. Some DBMS's provide two or three different types of user interfaces and therefore provide better flexibility for plant level operations.

PROGRAM DEVELOPMENT

The database management systems at the process level typically provide only a set of library calls to be issued by system management applications. These generally are visible only to the control system developers.

Plant level database systems however provide different database application development support tools such as Fourth Generation languages (4GLs), host language (e.g., Fortran, C, Pascal) -embedded SQL and automatic application generators. These tools reduce the development costs and times involved in the growing number of plant level applications.

Enterprise level database systems on the other hand have traditionally come with a number of mature, screen-based application development aids, report writers and transaction development facilities. Current DBMS vendors, to distinguish themselves will continue to provide more programmer-friendly, intelligent, and graphically oriented application design tools and integrate the DBMS with a wide variety of third-party software development / engineering [CASE] tools and other tools such as spreadsheets and report writers.

Since a basic reason for moving towards the commercial DBMS's in process industries is to increase the visibility of data and thereby promote innovative applications of that data, program development facilities provided by the DBMS's will be an important criterion in selecting a plant level DBMS.

DATA DEFINITION LANGUAGE

Process databases are defined and configured generally as a part of the configuration management activity involved in the operation of a control system. The configuration tool in the control system provides menu-based mechanism for defining the structure of the process data as well as the default values associated with some of the database attributes. The data locations, names and instances of other attributes are then configured. The process data definitions can typically be only changed during reconfiguration time.

Plant databases on other hand need more flexible database definition and configuration facilities. Many of the current relational DBMS products provide such flexibility using different methods for data definition using SQL or Forms or other user interface mechanisms. Data definition using enterprise database systems however is more batch oriented, that is, database scheme definition is done once and changed only periodically at controlled re-organization points in time. Graphical-oriented database configuration mechanisms are becoming available in some DBMS products. These mechanisms will ease the configuration of control systems greatly and as a consequence will improve the use of databases. Data definition is not totally a technical issue only however; it involves the organizational issues of control, ownership and maintenance of databases as a part of the global information resource management policies. Only a few of the DBMS products currently provide any tools for dealing with such organizational policy and procedure issues.

DATA ACCESS DIRECTORY

The data definitions are stored in a data dictionary which itself is part of the database and thus can be changed and manipulated easily.

DATA MANIPULATION SUPPORT

To manipulate process data, the database systems at this level provide a limited set of primitives which read /write selected attributes of a data object. These limited primitives are not adequate to provide the variety of data manipulations that need to be performed on plant level data. A rich set of data manipulation primitives for selecting data satisfying a variety of conditions, merging the selected data, grouping the data in different ways and converting the data into various user-oriented formats is needed at the plant levels. Most plant level DBMS's currently provide a powerful SQL as the data manipulation language. SQL provides a rich set of data manipulations. Some DBMS's also provide other types of data access and manipulation capabilities, e.g., an icon-based data selection specification. Non-relational DBMS's used at the enterprise level provide more procedural data access languages leaving the more powerful data manipulation capabilities to application programming languages. Some of the current DBMS's provide extensions to SQL to provide application-environment specific features; such extensions are needed for selecting a future plant level DBMS.

CONCURRENT ACCESS MANAGEMENT

Process database systems manage the execution of concurrent requests for data by maintaining data access relationships between data owners and data users which ensure that only one user is responsible for reading /writing that data and that user is also responsible for providing the data to any requestor. This method reduces the concurrency control overhead and ensures integrity of data.

Most of the DBMS products used at the plant level are multiuser systems and, as such, they provide an elaborate concurrency control mechanism to ensure the consistency of data while executing concurrent user transactions. Many relational DBMS's use locking for controlling concurrency while only a few provide referential integrity, i.e, the ability to keep all related data updated when a single piece is updated. Referential integrity as well as the performance and flexibility of the concurrency management mechanisms will be two key factors in selecting future plant DBMS's.

Enterprise database systems provide very efficient and complex transaction processing support mechanisms which control concurrency.

NETWORK ACCESSIBILITY

Process databases are naturally distributed throughout the control system (see below). To access that data, process database systems provide transparency of location by means of a symbolic name to an internal ID directory distributed across all of the nodes of the control system. Currently, most plant databases are typically centralized. However, the natural tendency occurring with the distribution of plant management operations is that of an increasingly distributed storage and management of plant data.

Some of the current plant level DBMS products support distribution and location independent management of data using the same DBMS product on the different nodes in the network. Only a few of these products currently provide the mechanisms necessary for ensuring the full integrity and recovery of the distributed data. It is important to consider if a distributed DBMS provides these recovery and integrity features that are flexible enough so as not to cause any severe performance penalties.

Since management of the operation of a distributed DBMS can lead to complex organization operational issues, it may be useful in some organizations to consider the flexible and performant distributed DBMS products becoming available recently. These products manage data on a central server but provide transparent access to that data using front ends which are distributed across the plant-wide network. They also support the use of distributed data on a flexible configuration of micros and minis on the network.

DATA FILE COMPATIBILITY

Process databases are typically in proprietary file structures and thus require custom interfacing software. Plant DBMS products typically can accept or produce data in a variety of standard formats such as ASCII and binary. Some plant DBMS products allow importing from/exporting to a standard set of application/user tools such as spreadsheets and report writers.

To transport data in a distributed DBMS, currently, standards organizations are developing SQL based presentation layer standards which will need to be considered in selecting a plant DBMS in the future. The necessity for data gateways between heterogeneous plant DBMS's also needs be considered in selecting the different systems at the plant level.

SECURITY

Security in process database systems is maintained at the user interface level. Since access to control systems is generally limited, this type of security control is adequate. However, at the plant level, more stringent security control is needed to restrict access to plant level data which is available typically on common computer platforms.

Current DBMS products at the plant level allow read/write privileges on different attributes of different data objects to certain users during certain times and under certain conditions. Only a few DBMS products provide mandatory security enforcement using such techniques as allowing users with right security level to access only the data at that security level. As more DBMS's provide such levels of security and as more users/applications access plant level data, security will become a key consideration in the selection of a plant level DBMS.

CONTROL SYSTEM INTEGRATION

In applying commercially available DBMS's to the support of real-time control systems, a balance must be struck between the ease-of-access from the point of view of all users and the time constraints of the real-time data collection system. The following are some key characteristics to be evaluated in selecting a DBMS for integration with a real-time plant control system.

DIRECT ACCESS TO FILES

The interface to the data collection nodes or system should be specified so as to provide a minimum of loading or interference to the operation of the data collector. A DBMS which supports direct access to files, requiring only a definition of fields and extents, is best suited to the broad needs of database integration in the plant-wide system.

RANGE OF DATA TYPES

The various arrays, enumerations and status words found in a process control system must be supported as well as the conventional real and integer types. The DBMS will require a multiplicity of those types in order to interface the diversity of types found in the different control systems.

DATA OWNERSHIP

The safety of the plant, as well as the efficacy of plant operations and the accuracy of accounting and the quality of the product, all depend on the validity of the data used in the control and management of the plant. The concept of the "Data Owner" as the keeper of a data element and the process which is solely responsible for the validity of the data is well established [19, 82]. In providing access to the data in the process-connected control and process management systems, this key to integrity must not be compromised. With the widespread implementation of high speed data communications networks, copying of data into other database fields and replication of data tables into user databases can no longer be justified to speed up access time. Direct access to the data owner's values provides the most timely and accurate data.

ENVIRONMENT SPACE

The overhead imposed on processing as well as the RAM and bulk memory required are significant considerations in applying a DBMS to a real-time control system. The facilities and space consumed should be minimized.

OPERATOR INTERFACE

Making effective use of the data managed by the DBMS to control the process implies the use of a user-friendly but efficient query language with ready access to all the needed information. On plant-wide systems this calls for an interface to the process database that meets those criteria.

APPLICATION DEVELOPMENT TOOLS

The simplification of decision-making procedures is an important part of optimizing the operation of the control system. High level language tools provided with the DBMS are needed to permit the development of repeatable routines for data selection and collection for use in analysis and decision-making.

HETEROGENEOUS NETWORK SUPPORT

An inherent part of integrating the plant information and control system is providing for the exchange of data between the different data acquisition, analysis and control systems provided by different vendors and operating on different hardware and software platforms. These diverse sources are selected for their optimization of their functions. Optimizing the application of the data provided by those systems is a key to the overall optimization of the operation of the process and plant.

REAL-TIME PERFORMANCE

Underlying all of the needs in process and plant control and information systems is the need for timely and accurate data. Accuracy and timeliness are interdependent. In applying a DBMS in a plant control environment, the balance between the differing needs of the users of the information and the timeliness of the result must be maintained.

Current database system products which offer relational structures with SQL capabilities continue to make advances in the performance and distributed data management features. Only limited instances of the use of these systems in real-time applications exist today. The use of these systems in soft real-time applications will grow as the performance of these systems increases to the level at which it is satisfactory for the real-time manipulation needs of centralized or distributed data.

However, for supporting the data management needs of restrictive soft real-time or hard real-time applications, these products must be redesigned with such features as:

1. Deadline-based query/transaction scheduling mechanisms;

2. Non-blocking concurrency control methods; and,

3. Main memory-based fast recovery algorithms.

Until such real-time database system products are available on the market, the choices left for automation system (process level use) or other real-time system implementors is: 1) to wait, 2) to use the best of the current DBMS products where the real-time constraints are soft and the product performance meets the worst case needs or 3) to design/implement a real-time database system dedicated for the needs of the system being built.

PLANT-WIDE INTEGRATION

Integration among process, plant and enterprise databases is quickly becoming a serious concern in many process industries. The interface which exports process data into a plant level database and vice versa is the first concern in this regard. Transformation of the structure and contents of process/plant data is an essential operation of this interface. Definition of a common interface language can improve the current ad hoc process-to-plant data interface development process.

Integration of the distributed plant data is the second concern. A data dictionary documenting the distributed data and the mappings among that data and a methodology for maintaining an active data dictionary are important in plant level distributed data integration. Also, when heterogeneous DBMS's are used at the plant level, tools which ease the development of data gateways between these different DBMS's need to be considered. Export of plant data into the enterprise database and import of enterprise data into the plant is the third concern. Simple file-based export / import mechanisms are adequate in achieving this type of data integration because plant-to-enterprise data is exchanged less frequently than process-to-plant or across-plant data.

In considering the overall functions of DBMS's to be implemented in a plant, there are several aspects which are key to the effective control of plant operation. They may be described here in relation to the access the systems provide: to the process; to the operator; from the Plant Information Network; and, from the Enterprise.

ACCESS TO THE PROCESS

Timely acquisition of data from the process and providing access to that data is the prime concern of a control system. Meeting those needs while maintaining access to other users of the data is its principal function. Support for higher level control processors with timely access to the data is implicit in this. Processing of highly structured queries is a secondary function and must be kept from compromising the primary goal.

ACCESS BY THE OPERATOR

Effective access by plant operations to the process and all that affects it is the basic objective of the control system and the database which supports it. Providing access to the process data for analysis and decisions and for control to implement plant management policies is its reason for existence. It must continue to present the needed data to the operator in a timely, accurate and comprehensible manner.

ACCESS FROM THE PLANT INFORMATION NETWORK

In today's high technology operating environment, much of the process control data manipulation has been mechanized. Data from the process is readily available at the operator's fingertips on a CRT. The data from the rest of the plant and from the business which affects his decisions is also mechanized and may also be available through CRT's. The most effective use of this latter information is to also put it at the operator's fingertips. By juxtaposing business data with affected data from the process, more timely, balanced decisions can be made in the operation of the process. This would be similar to displaying the setpoint and output values of a controller loop alongside the process variable value as has been standard practice for years.

ACCESS FROM THE ENTERPRISE

As part of the plant management structure, the results of the operators analysis and decisions must be reported up the chain-of-command. The collection, analysis and decision-making at the higher levels requires those inputs. Likewise, the passing down of operating policies based on the higher-level decisions completes the plant level control cycle. For timely and effective flow of this information to occur, the databases in each organization must be integrated.

Often the information must flow between the databases of different makers over networks with different protocols. While MAP/TOP and the ISO protocols offer a long-term solution to the network integration problem, they do not address the database integration problems. Also, the networks available for short-to-mid-term implementation do not provide complete data integration since no migration plan is included in their specifications. The integration of this data from diverse sources for use throughout the plant is the principal challenge for plant information and control system development in the next decade.

DISTRIBUTED DATABASES IN THE FACTORY

In the discussion to follow, a large number of specialized databases will be described. It must be borne in mind that these are all parts of the overall global database and must obey all rules of variable naming, data dictionary and data directory, access methods, etc., necessary for their effective utilization as such. The apparent segmentation described below is for conveniences of description, of task use or of geography and must not be allowed to invalidate any part of the global database and global access in the system [79].

THE PROCESS SENSOR DATABASE (LEVEL 1)

Today most sensor signals are analog which are then converted to digital values at the controller or data acquisition unit. With the advent of digital sensor loops, more information becomes available across the wire than simply the process value and the loop continuity status [79]. All the variable parameters of the sensor become adjustable and readable. This set of values will come to constitute the loop database. While these values have been traditionally the concern of the instrument maintenance technicians, their digital form lends itself to making them available and adjustable by people working at the higher levels of the hierarchy. Figure 7-1 presents the stages of sensor data reduction important in considering the process sensor database.

Process engineers at the process management level and operators at the process-connected level may now access sensor and actuator parameters for monitoring of scales, range and calibration. Sensor adjustments are more easily made and better record keeping of changes is practical with an on-line sensor database. Maintenance can be scheduled according to trends in the sensor parameters. This elementary data set becomes the first level database in the plant information system hierarchy.

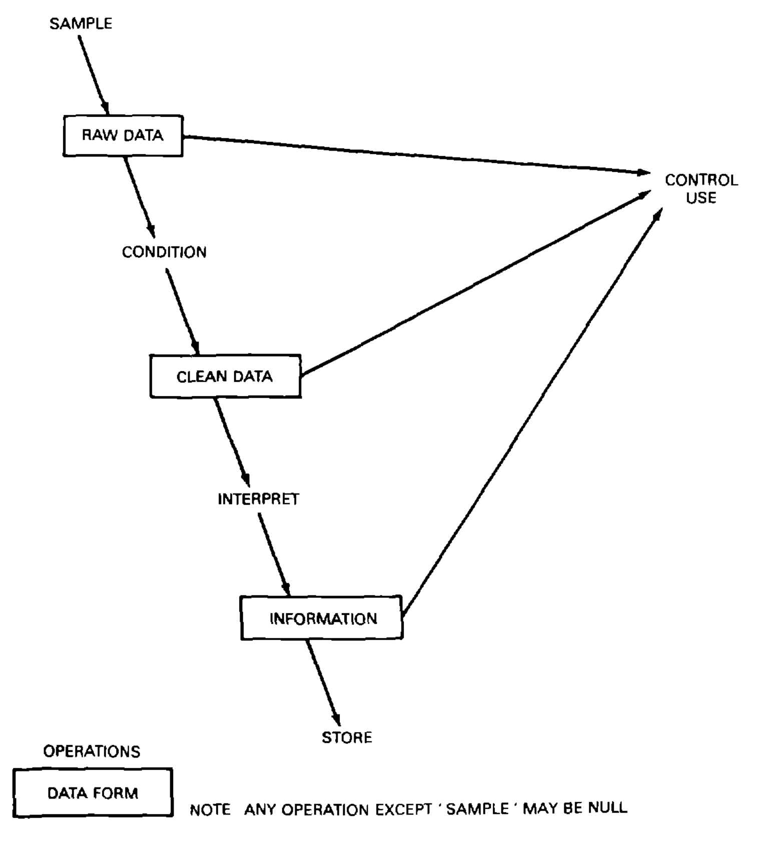

Figure 7-1 What do we do with data?

Digital control at the process-connected level has been well established over the last decade. Many functions, which formerly required a central process computer, have migrated out to controllers operated by microprocessors at this level. Batch process sequencing and control as well as sequential logic control have become integrated at this level. This database contains the loops' control and status values expressed in forms which are suitable for processing at rates in the 1 second class.

In the distributed database on the data highway, the *Loop algorithm* values support regulatory control. Sequential control as well as batch sequencing are supported by the *sequence state* database. Also, safety control depends on the *interlock logic* database. Similarly, analysis of process trends is provided by the database of *process snapshot* values. Today the database associated with process-connected devices is generally the basis for higher level databases.

THE PROCESS CONTROL DATABASE (LEVEL 2)

A process control database, called the *Unit Database* in Figure 6-15, is required for Level 2 operations in the control of each process unit of the plant. McCarthy [79] has presented the following discussion of these databases.

The distributed database at the control room must be integrated so as to meet the needs of the operations management tasks performed there. Process operation and management which occurs at this level and the higher level control strategies which are implemented here require a more sophisticated database than the lower levels. The distributed database must be partitioned and structured to meet user needs which vary across a broad range. A discussion of this partitioning and the structure to achieve it follows.

The *Process Variables and Attributes* database provides access to the values in physical (engineering) units representation. The data obtained from the process units and sensors as well as the values derived from them are accessible with greater precision for extensive calculations and transformation into control indices. These values become the "Global" Process Database where information can be accessed by Tag and Process Variable or Parameter Names independent of physical locations.

Thus the Process Control Database has a dual expression. It is on the one hand globally accessible as above and on the other hand is defined as data structures representing particular generic application functions as described below.

The *Graphics Image* database supports the display interface to the operator, simplifying interpretation of operational situations. Access to the picture elements and abstracts data enables the operator workstation to display the process information in a meaningful fashion, thus simplifying and expediting the analysis by the operator.

The *Control Algorithms* and *Computation Language* and the associated process databases permit the implementation of cross-unit and plant-wide control strategies. The language processors executing the control functions have access to a wide variety of procedures and functions in addition to the variables necessary to complete their tasks.

Alarm management is achieved with the *Event Sequence* database. The events occurring at the process as well as those interpreted from the data and calculations are organized, prioritized, and reported for use by the operator workstation and other Data Users.

Trends in process conditions and states are determined from the *Process History* database. Data collection and storage in timed snapshots and averages permits the later analysis of historical values.

Sequence-of-events analyses and reports are supported by the *Event History* database. Events occurring in the process and in the control room are time-tagged and entered into journals for later analysis.

Control system maintenance and management is sustained by the *System Event* database The system error messages are time-tagged and journalized for later diagnosis and predictive maintenance action.

The generation of logs and reports is provided for by the *Format Generation* database. Standardized reporting forms can be accessed from throughout the system, thus easing their availability and use.

As the entry level of the corporate decision-making chain, the operator requires access to the databases at all levels of the organization. The control room database is structured to simplify and expedite decisions. In addition, the needs of other organizational groups for information from this area and the relaying of operating guidelines to this area are accommodated.

DATABASES FOR HIGHER LEVEL FUNCTIONS

It is when the various department databases which were formerly required are interconnected via the data highway network that the challenge of integration becomes most apparent. These databases originally grew up in card files, file cabinets and typed and hand-written reports. When this information handling was mechanized, the database system and machine selection were guided by the inherent structure of the data (and the media). Now that a plant-wide data communications system is becoming available, the need for these diverse systems and data structures to interact with each other must be addressed. And, they vary over a considerable range.

The Quality Control laboratory and statistical functions, the production control planning and scheduling, as well as the Process Control Engineers' modeling and analysis each require their own database in addition to access to others to complete their tasks. In a like manner, Maintenance Management, Building Management, Materials Management, Order Entry and Tracking as well as Manufacturing Requirements Planning establish their own databases and require access to and from others.

The integrity of information or data values becomes a concern when databases are linked. The security of each database is part of this concern. For effective use, each user must be confident that the data received from each query is valid. The concept of Data Owner and Data User introduced in the distributed database of the Process Control Network provides this mechanism. There the process in the node which produces the information is the owner of that variable. Only that process owner can change its value. All attempts to input variables to the process Data Owner are checked for validity of access. The distributed

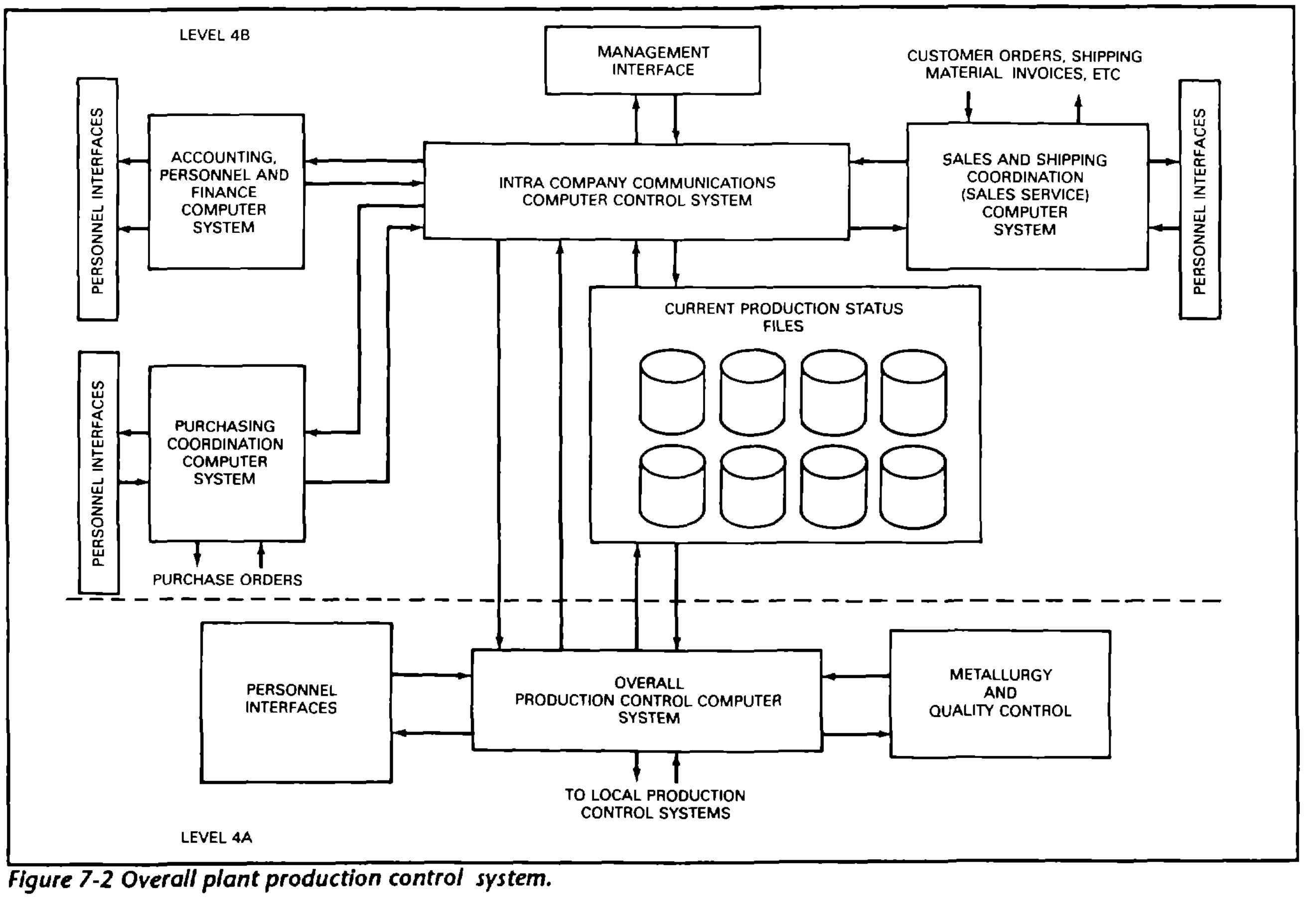

Figure 7-2 Overall plant production control system.

database manager in each node assures this in its response to data dictionary queries. The Data Owner-Data User concept can be extended to all levels. Table 7-1 presents a set of important concepts concerning Data Ownership important to the integrity of plant databases.

The need for timely and accurate data for decisions raises another concern. That is the duplication of data values across the plant-wide database. Up until now departmental database systems have routinely collected data held by other Data Owners and entered them into their structure. This was primarily done to overcome the delays in acquisition due to the latency of the slow speed networks available previously. The penalty was the loss of timeliness of the value. The high speed of today's networks removes the need to acquire the value in advance, thus removing the need for duplication. It eliminates the possible inconsistency between the actual value held by the Data Owner and the transient value held in the duplicating database.

The departmental databases located throughout the plant all come together on the network. Gateways to higher and lower levels interconnect those databases with the entire plant and corporate structure. MAP provides the tools for intercommunication. The distributed database manager coordinates the interaction.

Figure 7-2 presents several concepts concerning the centralized portion of the overall database of the plant production computer control system. Figures 7-2 and 7-3 [90] show this database labelled as the Current Production Status File. Figure 7-2 indicates that this file is maintained by the Overall Production Control Computer System of Level 4A. Management and the staff functions of the company have a READ access to this file but not a WRITE access. The latter is restricted to the Production Control System. Figure 7-3 presents a diagrammatical presentation of the proposed contents of this file and the association of each element with the staff function most likely to use this data.

TABLE 7-I

SOME DATA MANAGEMENT CONCEPTS IMPORTANT TO THE DESIGN OF THE CIM INFORMATION MANAGEMENT AND AUTOMATION SYSTEM CONFIGURATION

1. ALL PLANT COMPUTER DATA FUNCTION IMPLEMENTATIONS WILL ACT IN THE ROLE OF "HELPFUL" HUMANS IN RESPECT TO TRANSMISSION OF ESSENTIAL INFORMATION.

   DATA OWNERSHIP COMPRISES THE RESPONSIBILITY TO PROVIDE AND MAINTAIN THE INTEGRITY OF THE IMPLEMENTED SYSTEM POLICY IN RESPECT TO THE INFORMATION CONTAINED IN THAT DATA. IT *CANNOT* IMPLY ANY CHOICE OR RESTRICTION OF DISTRIBUTION OF INFORMATION ESSENTIAL TO ANOTHER ENTITY.

2. FOR EACH DATA ITEM IN THE SYSTEM, THERE IS SOME SINGLE ENTITY THAT IS RESPONSIBLE FOR THAT ITEM. ALL DATA ACCESS BETWEEN ENTITIES MUST BE EXPLICIT. DATA INTEGRITY REQUIRES EXPLICIT ACCESS FOR ALL WRITES. DATA SECURITY REQUIRES EXPLICIT ACCESS BOTH FOR

   READS AND WRITES. EXPLICIT ACCESS ALSO CONTROLS SYSTEM COMPLEXITY AND MAKES SYSTEMS EASIER TO MAINTAIN.

3. A SYSTEM-WIDE ACCESS METHOD MUST BE DEFINED. THE ACCESS METHOD WILL SUPPORT THE CONTROLLED INTERCHANGE OF DATA BETWEEN ENTITIES CONSISTENT WITH [2] ABOVE.

4. DATA PASSED BETWEEN ENTITIES IS OF THREE TYPES:

   A. RAW DATA, WHICH EMERGES DIRECTLY FROM A SENSOR,

   B. CLEAN DATA, WHICH HAS BEEN SMOOTHED, AND CONDITIONED

   C. INFORMATION, WHICH HAS BEEN ASSOCIATED WITH A SEMANTIC CONTEXT.

AS NOTED IN FIGURE 7-1.

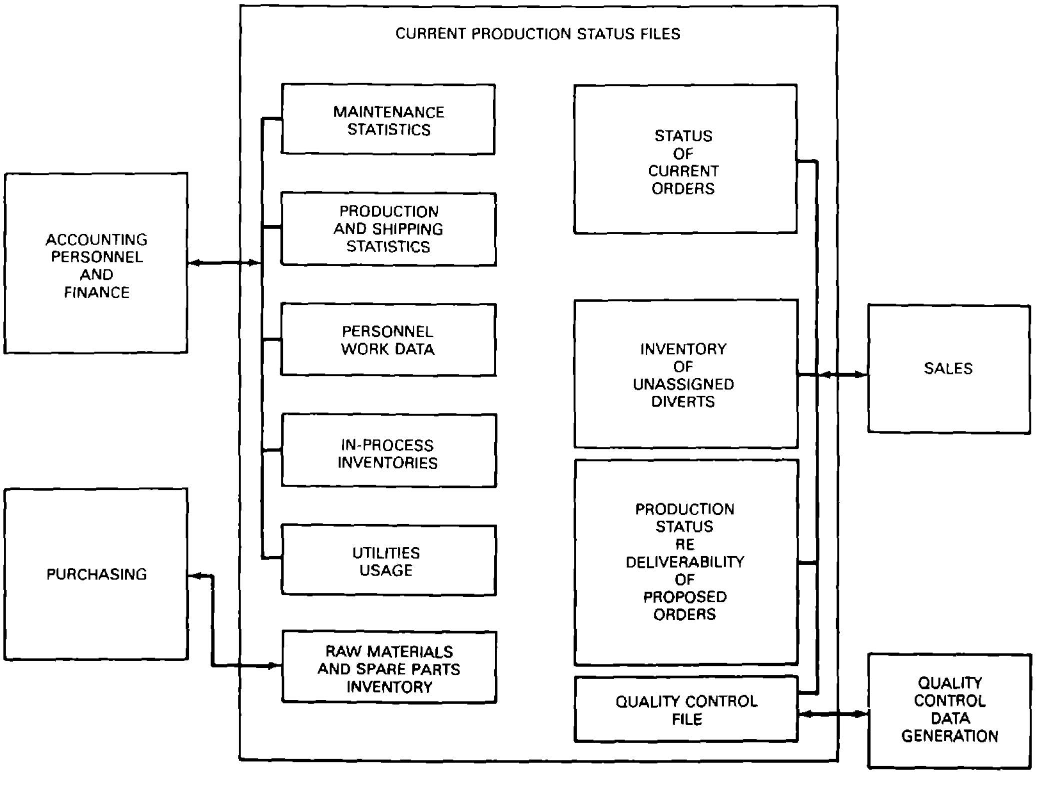

Figure 7-3 Non-operational contact with hierarchy computer control system (Level 4B), the production status file.