A Very Large Database System to Serve National Welfare

Michinao Mori and Kenji Suzuki
NTT Electrical Communication Laboratories

Hideo Abe and Kenji Itoh
NTT Data Communication Bureau
JAPAN

ABSTRACT

This paper describes the present state of construction of a very large social insurance database system. This large-capacity, high-traffic-volume database system has real-time processing capabilities for handling enormous volumes of data that encompass the entire nation.

1. INTRODUCTON

The data communications systems offered by NTT as part of the company's telecommunications services have tended to become increasingly larger in scale in recent years (Fig.1)[1]. This is particularly true of those systems set up as national projects on a nation-wide scale, which store enormous volumes of data ranging from several tens to several hundreds of gigabytes. In the past five years, both the volumes of data and processing handled by these systems have expanded twofold.

One of the largest of these national project systems is the social insurance system, a very large capacity, high-traffic-volume database system. This system handles data records encompassing the entire nation and is capable of processing tremendous volumes of data in real time.

This social insurance database system includes four individual programs under the jurisdiction of the Social Insurance Agency in the Ministry of Health and Welfare: health insurance, national pensions, welfare pension insurance and seamen's insurance. The system was designed to improve and expand social insurance services by upgrading the level of administrative service and by enhancing the efficiency and speed at which operations are carried out.

The number of people covered under the social insurance services of these four programs includes some 68 million people who are currently insured and approximately 13 million people who are entitled to receive pension benefits (as of March 1984). The tasks that must be processed in the course of providing these services include the entering, updating and deleting of the records of the people insured, from the time they enter one or more of the systems until they begin to receive benefits in old age; in addition, collecting insurance premiums and paying pension benefits must also be processed. The social insurance database system thus has to handle an enormous volume of complex processing tasks involving data records that extend over long periods of time. Since it must store and manage large volumes of data for such extended periods of time, it has been desiged with a gigantic database capacity that will eventually exceed 200 gigabytes.

In designing and constructing this system, one of the major issues that had to be dealt with was how to configure and operate such a mammoth database so as to provide maximum efficiency in retrieving and manipulating data.

The services offered by the system and the tasks it handles are outlined in Table 1, while the essential requirements imposed on the database are shown in Table 2.

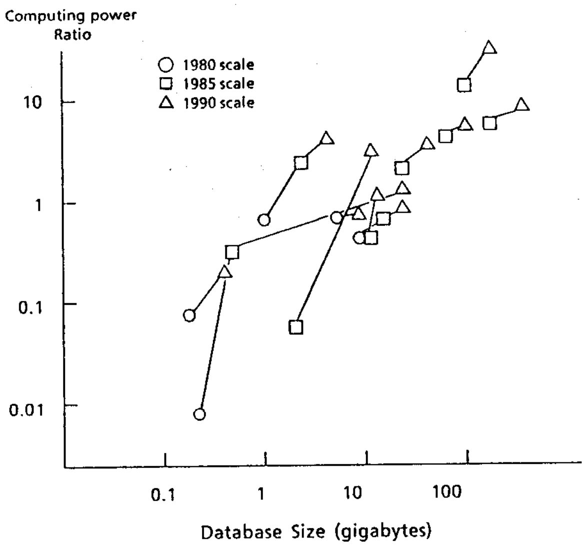

Fig.1 Trends in database system scale for NTT's data communication services [1].

Proceedings of the Twelfth International Conference on Very Large Data Bases **Kyoto, August, 1986**

2. SYSTEM CONFIGURATION

The system has a functionally distributed configuration with dedicated database management processors (BEPs) that are independent of on-line and off-line host processors which perform the processing operations. This distributed configuration allows easy use any database within the system as if there were a single database as well as highly efficient processing of the complex and voluminous tasks involved in social insurance services.

The BEPs are composed of two tightly coupled multiprocessors to assure high reliability and sufficient processing capacity. The host processors have a duplex configuration with some multiprocessors aspects.

As for the links between host processors and BEPs, CPU multiconnectors are used to connect intra-center processors and leased lines are employed to connect inter-center processors. The leased lines connecting processors have a capacity of 48kb/s, while the computer center and social insurance offices are linked by 4800 b/s lines. Social insurance offices are equipped with kanji window machines and kanji printers.

The database is made up of 400-megabyte magnetic disks as the storage medium, and it is expected that eventually there will be approximately 1,000 spindles.

The system configuration is illustrated in Fig.2.

3. DATABASE CONFIGURATION

The database has a distributed configuration in which it is divided and allocated among multiple BEPs. The application program being run on a host processor can access data stored anywhere without being aware of the location of the database. Each processor is provided with the same distributed database management system (DEIMS-3)[2] so as to achieve a complexity homogeneous database system.

DEIMS-3 is a general purpose database management system (DBMS) that runs on NTT"s standard computer system DIPS. In addition to providing database management functions based on the CODASYL network model[3,4], DEIMS-3 also offers distributed database management functions employing the database access protocol (DBAP) which has been proposed for accessing heterogeneous databases uniformly in data communications network architecture (DCNA)[5].

3.1 Schema configuration and database allocation

The distributed database management functions provided by DEIMS-3 are categorized into four hierarchical schemas: local internal, local conceptual, global conceptual and global external.

The local internal schema consists of a logical schema, storage schema, physical schema and hige-level schema. The logical schema describes the data structure in the same manner as the CODASYL network model. The storage schema describes the method of allocating data, indexing procedure and method for setting the pointer. The physical schema describes the mapping of data to the memory storage space managed by the operating system The high-level schema is a description of an application data view and it describes the next local conceptual schema in detail.

The local conceptual schema is a description that virtualizes a local internal schema to the network; it employs the DBAP data model (VDB) of the DCNA. The VDB model is a integrated data model that includes the representation components of hierarchical, network and relational models. At present DEIMS-3 incorporates the hierarchical and relational data models, which are called a hige-level hierarchical model and the schema is called a high-level schema. When DEIMS-3 is used to construct a homogeneous distributed database, the local conceptual schema is not necessary since it is replaced by the high-level schema of the local internal schema.

The global conceptual schema describes a complete view of the entire system, integrating the local conceptual schema and it employs the VDB model.

The global external schema is a high-level hierarchical model that describes an application program's view of the database. A high-level database manipulation language (HDML) is provided to handle requests for database manipulations from an application program to the global external schema. This helps to reduce the communication load in the distributed processing configuration.

In the social insurance database system, the global conceptual schema and global external schema are allocated to the host processors, while the local internal schema and local conceptual schema are allocated to the BEPs. Various data storage arrangements are used between the BEPs to allow data to be divided among different communities and to assure high access efficiency. These include divided, redundant and independent data storage, but data are apparently allocated to be divided without any redundancy by means of areas described in the schema. Areas are the basic unit of logical storage in the database and are subdivided into pages and records. An application program on a host processor can access the database by specifying one of the entry keys that uniquely identify the record occurrences forming the hierarchies. The global conceptual schema manages the information that provide the correspondence between the area names or entry keys and the relevant database nodes (in this case BEPs). Consequently, the distributed organization of the database is invisible to the application program. An example of the arrangement between the schema and the database is illustrated in Fig.3.

The data structure used in this database system consists of 27 different hierarchies. The data structure pertaining to people currently insured under the four programs involves the largest number of different types of records, 57 altogether.

3.2 Database access

Access to the database on a BEP is performed in parallel between the on-line host and off-line host that performs batch processing during the daytime. Access to a database following batch processing normally becomes a long transaction. However, in this system, access operations are divided into short transaction and executed, just as in on-line processing. This feature provides greater parallelism with short-transactions in on-line processing.

Competition for access to a database on a BEP can cause deadlocks. In this system, local deadlocks within a BEP are prevented when DEIMS-3 locks the requested record. Global deadlocks between a host and a BEP are detected by a time monitor included in the distributed control functions of DEIMS-3.

3.3 Database backup

Since the volumes of data stored in this database system are so large, a database journal for updating the database is first collected in the magnetic disk unit and a modifications journal which is edited database jounals is collected in magnetic tapes. This approach greatly eliminates much of the labor involved in magnetic tape operations.

In addition, instead of executing total dumpings of such large volumes of data, an integrated system has been adopted whereby cumulative modifications can be run during the daytime in order to shorten recovery time after failure. Through cumulative processing everyday of the modifications journal, backup magnetic tapes are created that contain the same contents as if total dumps of data accumulated on magnetic tapes were executed at regular intervals.

3.4 Database reorganization

The working time required to reorganize a database becomes more of a problem as the volume of stored data grows larger. In this system, the database may be reorganized at regular intervals, sporadically or in responce to some event that has occurred. The first type of reorganization is carried out according to a planned schedule before the estimated interval for reorganization elapses and available space in the database is filled. Sporadic reorganization is executed when surplus space has been filled by the addition of unexpected data before the anticipated interval for reorganization has elapsed. The third type of reorganization is carried out in order to cope with some event such as the establishment of new offices or the division or integration of existing offices.

Examples of the database reorganizations implemented at regular intervals to date are shown in Table 3.

4. CURRENT STATUS

The first version of the social insurance database system was put into service in January 1980. This paper has described the second version of this very large database system which was inaugurated in February 1984. Work is now proceeding on the completion of those tasks which have yet to be incorporated into the system.

ACKNOWLEDGEMENTS

The authors would like to thank the Social Insurane Agency in the Ministry of Health and Welfare for giving them an opportunity to carry out this development.

REFERENCES

1.Suzuki,K., Tanaka,T., and Hattori,F. Implementation of a Distributed Database Management System for Very Large Real-time Applications. In *Proceedings of IEEE COMPCON Fall*(1982),569-577.

2.Mori,M., Yoshida,K., and Suzuki,K. Development of a Distributed Database Management System. *Review of the E.C.L., NTT. 33,3* (1985),443-449.

3.CODASYL Data Base Language Task Group. CODASYL COBOL Data Base Facility Proposal (Jan 1972).

4.CODASYL Data Description Language Committee. Data Description Language Journal of Development (1978).

5.Kawazu,S.,et al. DCNA Database Access Protocol. *Review of the E.C.L., NTT. 30,1* (1982),197-212.

Table 1 Applied Services and Tasks

Tasks		Programs				Contents
		Health Insurance	Welfare Pension Insurance	National Pension	Seamen's Insurance	
Short-term Tasks	Insurance Application	CS	CS	CS	FS	Enter the records of the people insured, make out the pension pocketbook and the insured certificate, etc.
	Premium Collection	CS	CS	CS	FS	Investigate, decide, calculate, notice and receive the premium.
	benefits	FS	---	FS	FS	Pay the short-term benefits.
Long-term Task	benefits	FS	FS	FS	FS	Decide and pay the pension.
Inquiry		---	CS	CS	CS	Inquire the qualification and recipient record, etc.

<NOTE> CS : Current Services offered by this system
FS : Future services
--- : None

Table2 Database Requirement (1986.3)

Database Size :

Total Volumes : 533 Spindles

Main Databases :

Database Names	Average Records Length(bytes)	Total Number of Records
Name Index of National Pension	128	54,000,000
The Forfeit People Insured of National Pension	174	51,000,000
The Current People Insured of National Pension	477	28,000,000
The Forfeit People Insured of Welfare Pension Insurance	209	94,000,000

Traffic Volumes :

Online Transaction	162,000	transactions / hour during peak period
Offline Transaction	500,000	transactions / hour during peak period (Correspoding to 500,000 people insured)
Total	662,000	transactions / hour during peak period
Number of Database I/Os :	800,000	times / hour during peak period

Journal Volumes :

Average : 2,000,000 blocks / day (1 block = 4096 bytes)

Number of used DASDs : 30 spindles / day

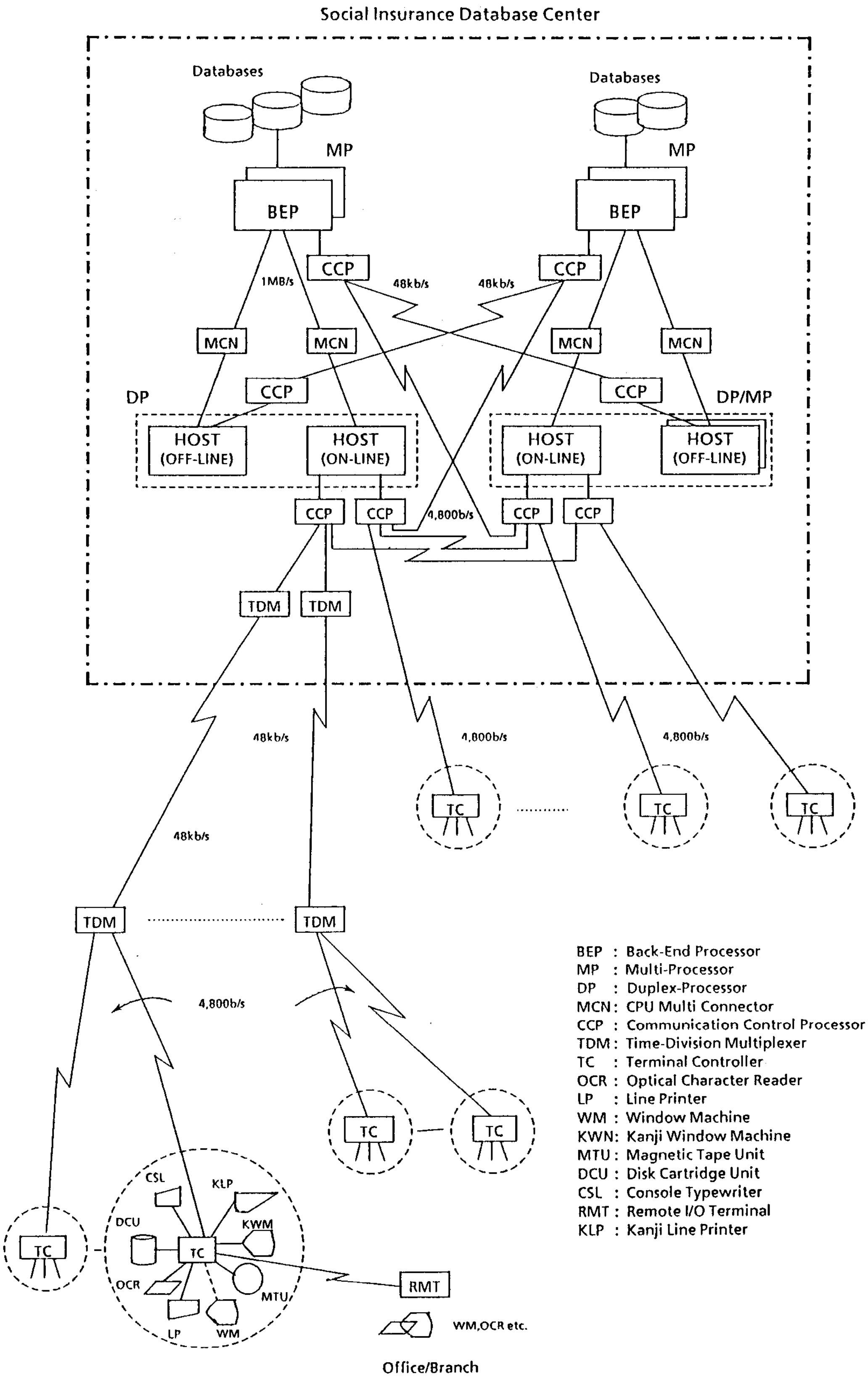

Fig2 Social Insurance System Organization

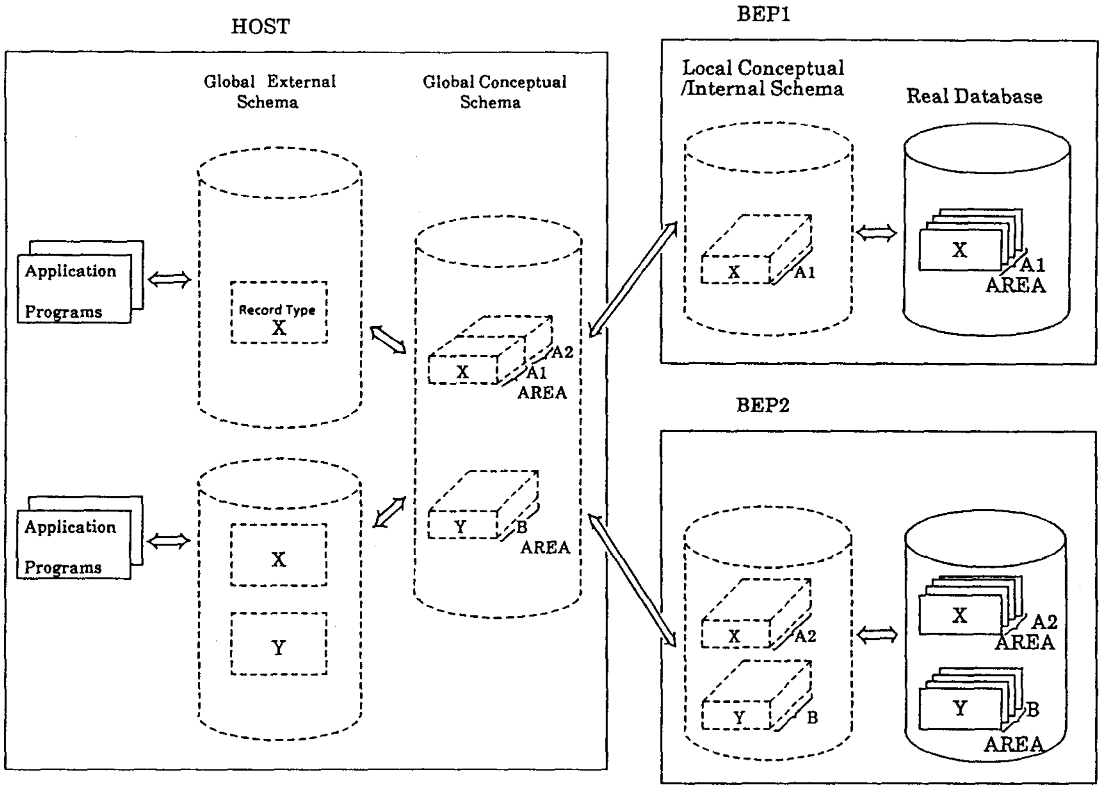

Fig3 Schema and Database Allocation of Social Insurance System

Table3 Example of the Database Reorganization Implemented at Regular Intervals to Date

Databases	Number of		Total Elapse Time (hour)
	Objective Realms of Reorganization	New Realms after Reorganization	
Name Index	33	45	41
Office	21	25	30
The Current People Insured	128	170	110
The Forfeit People Insured	59	74	63