

ANALYZING TRANSACTION ACTIVITY  
TO A LARGE DATA BASE: AN EMPIRICAL STUDY

BY

RICHARD D. HACKATHORN

UNIVERSITY OF CALIFORNIA

IRVINE, CALIFORNIA 92664

Information systems that manage large data bases are becoming a major factor in the ability of organizations to cope with the demands of our society. An emerging problem with these information systems is that their overwhelming complexity breeds an inherent inflexibility to change. Hence, changes occur when forced by crises, and not in response to plans of graceful evolution.

This paper reports on the initial phase of work to develop a systems analysis technique that will assist in the evolution of an information system. The technique, called Data Base Activity Analysis, views the information flows within an organization from the vantage point of its data base. By analyzing the transactions to a data base in terms of what they do to entities within the data base, the analysis technique attempts to reconstruct the information system design, as reflected by the current usage of the system.

Property Data Base System

The work was conducted in the County of Riverside, California. The county government utilizes an IMS/2.3 data base management system operating on an IBM 370/158 computer with two million bytes of main storage and over two billion bytes of online disk storage. The focus was on the property taxation data base that has been operational for several years. The data base contains over 400,000 land parcels, requiring approximately 500 million bytes of disk storage. The data base is heavily transaction-oriented with roughly 12,000 transactions per day performed upon the data base. Over 35 remote terminals, each of which can utilize over 100 transaction programs (barring security restrictions), are devoted to maintaining the data base.

The five departments currently involved with the property system are: Assessor, Auditor, Tax Collector, Recorder, and Building Department. The functions performed by the property system are: parcel identification, assessment valuation, exemptions, value certification, tax rate establishment, tax accounting, and public service.

---

This work was supported in part by a grant from the Chancellor's Patent Fund for Graduate Student Research. Author's address: Graduate School of Administration, University of California, Irvine, California 92664.

### Collection of Transaction Data

Standard IMS system log tapes were processed during off-hours to extract data on transactions to the data base. The primary data items collected on each transaction were: (1) the time and date that the transaction was entered; (2) the action performed by the transaction; (3) the terminal entering the transaction; and especially (4) the identifier of the entity in the data base that was referenced by the transaction. The identifier was obtained by scanning the input message text of the transaction. Although other data items are available from the system log (such as DL/1 call counts, completion code, elapsed time, and priority), they were not utilized in the study.

The two types of IMS log records used are: (1) the input message prefix record; and (2) the application accounting record. By sorting on time, these two record types are combined into a single transaction record. A series of four COBOL programs sort the data on time, transaction code, terminal, and identifier and print appropriate listings of each sort. The last program also generates a file of transactions ordered on identifier for further analyses.

### Transaction Statistics

On a typical day, the County of Riverside processes a total of 18,000 to 20,000 transactions to the IMS data bases, of which 12,000 (60%) are to the property data base. Between 85% to 90% of the property transactions can be attributed to an identifier and, hence, to a particular land parcel in the county. When sorted on identifier, roughly 6,000 unique land parcels are listed. Hence, there is an average of two transactions to a land parcel during a day of operation. Typically, 60% of the land parcels have only one transaction to them; however, there are usually six or eight parcels having ten or more transactions. So far, the maximum has been 18 transactions to a land parcel during a day. About 70% of the land parcels have transactions with the same transaction code entered by the same terminal; however, several hundred land parcels (2%-3%) have transactions consisting of different transaction codes coming from different terminals.

### Benefits to County

Even though the results have been preliminary, several tangible benefits have results so far. First, the report listings provided the first detailed view of how users were actually inquiring and updating their data bases. For instance, a design error was discovered in one transaction program that did not allow a certain combination of parameters. At one time, the combination was deemed illegal; however, changes to other transactions negated that assumption. Clerical personnel were entering two transactions to "work around" the problem. Secondly, the listings provide a complete audit trail of every transaction (even errors) entered into the system. Thirdly, the listings have been used as a training aid to point out instances in which persons need training with certain transactions. Finally, a "frustration complex" problem was noted in which some users would strike the Enter Key on the terminal repeatedly to "speed the system up."

### Conclusions

Preliminary results based on the Data Base Activity Analysis work have produced several tangible benefits for the County of Riverside and some interesting statistics concerning transaction activity to a large data base. In particular, the data show a surprising degree of richness in terms of multiple transactions to a typical entity within a data base during a one-day period. Moreover, interactions among user departments concerned with the same entity occur frequently.

Current work is concentrating on detecting patterns of transaction activity within the information system and on relating those patterns to activities generally within the organization. Determining the context and frequency with which certain transaction sequences occur is important to balancing the considerations of efficiency versus flexibility in designing an information system. This and related research will hopefully develop the means of evolving easily an existing information system to serve better the needs of the organization.

### ACKNOWLEDGMENT

The author wishes to thank the County of Riverside and, especially, Mike McLoughlin and Dennis DeBolt of the Data Processing Department for their extensive cooperation in this study. The advice of my dissertation advisor, Professor Fred Tonge, has also been invaluable.