

A SYSTEM FOR
AUTOMATIC GENERATION OF RELATIONAL
DATA BASES

BY

MEIR COHEN

B.S., TEL AVIV UNIVERSITY, ISRAEL, 1979

A MASTER'S REPORT

submitted in partial fulfillment of the
requirements for the degree

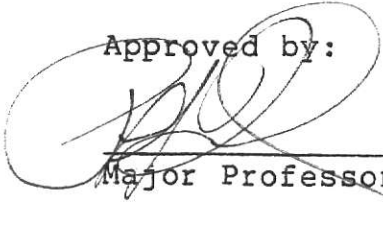
MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1981

Approved by:


Major Professor

SPEC
COLL
LD
2668
R4
1981
C63
c.2

TABLE OF CONTENTS

1. INTRODUCTION
2. DATA BASE DESIGN - THE METHODOLOGY
3. SYSTEM DESCRIPTION
4. AN EXAMPLE
5. CONCLUSION
6. APPENDIX A - COMMANDS' SYNTAX-GRAPHS
7. APPENDIX B - DETAILED CODE

Introduction

Data base management system (DBMS) technology has achieved acceptance and wide application. This is demonstrated by numerous surveys and the availability of several textbooks dealing exclusively with the topic. However the usefulness of a DBMS is directly related to the way in which the actual data base was designed. Numerous papers have been published dealing with the design phase of data base implementation in an organization. Some agreement exists among these researchers as to the categorization of the design process in two of these phases- the logical design and the physical design. However, agreement has yet to be reached about the overall design process.

Most researchers agree that design should be automated as much as possible. One of the basic components in the design is a descriptive mechanism that represents the information input and the results of each design step. Hollist (Hollist [80]) called this mechanism the Design DBMS.

Among the characteristics required in a Design DBMS, one can identify the ability to detail data names, types, domains and ranges. In addition the design DBMS should be easy to operate and the processes which it automated as possible. The commands should be flexible and in an English-like form.

This report deals with the implementation of the Document Handler, a tool in the logical design of data bases on the

Interdata 8/32 at Kansas State University. The methodology described in the second section is based mainly on the work of Fisher (Fisher [79]). The third section gives a detailed description of the structure and operation of the Document Handler. An extensive example is given in the fourth section. This demonstrates the capabilities of the Document Handler.

Data Base Design - The Methodology

Introduction

One can divide the design process into three general processes:

- 1) Organization survey and general design
- 2) Detailed design
- 3) Key Specification and Schema Translation

Although there seems to be a clear cut distinction between these processes, this is not the case. It will often be required to go back from process 2 to process 1 and from process 3 to process 2. In all these processes, the design DBMS is of great help.

The following sections will describe the design in more detail.

Organization Survey and General Design

The first thing designers ought to do is understand the structure of the organization, its operation, its problems and its aims. This can be accomplished by a series of lectures and interviews with the management of the organization under investigation.

When this understanding is reached, the designers can collect all the documents currently used by the organization. The documents and all the columns in them are entered into the design DBMS with a specification of their use as input, output or resident documents. Attributes such as type, range of values and security constraints can be associated to the columns.

At the conclusion of this step, the design DBMS will summarize and format the information so that it can be used in the next step in the design process.

Detailed Design

Three activities take place in the detailed design :

- 1) Deletion of synonym column names

- 2) Removal of insignificant columns
- 3) Solving undeclared output columns

These activities are not necessarily done in this order. The documents and the columns in them are scanned one by one and a decision is made as to the appropriate action to be taken.

It might be necessary to go over the information time and again in this step to ensure that the design is complete and that the integrity of the data is maintained. It is certainly better to repeat the process at this time and not later, when the cost would be enormous.

Deletion of Synonym Columns

In many cases, different names are used to specify the same entity or attribute of an entity in the organization. This is also reflected in the documents of that organization. For example, the name "DATE" in document A might refer to "SHIPPING DATE", as it appears in document B. It is apparent to the user of document A that this is the meaning of "DATE", these two names should be identical and hence "DATE" should be renamed "SHIPPING DATE".