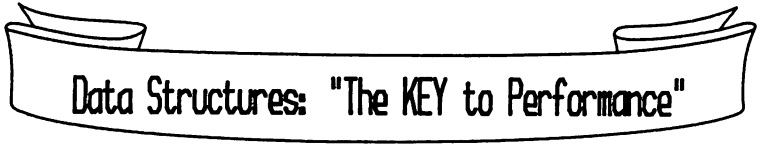


- a tutorial -

by David G. Robinson

PowerSpec

International

Data Structures: "The KEY to Performance"

Abstract

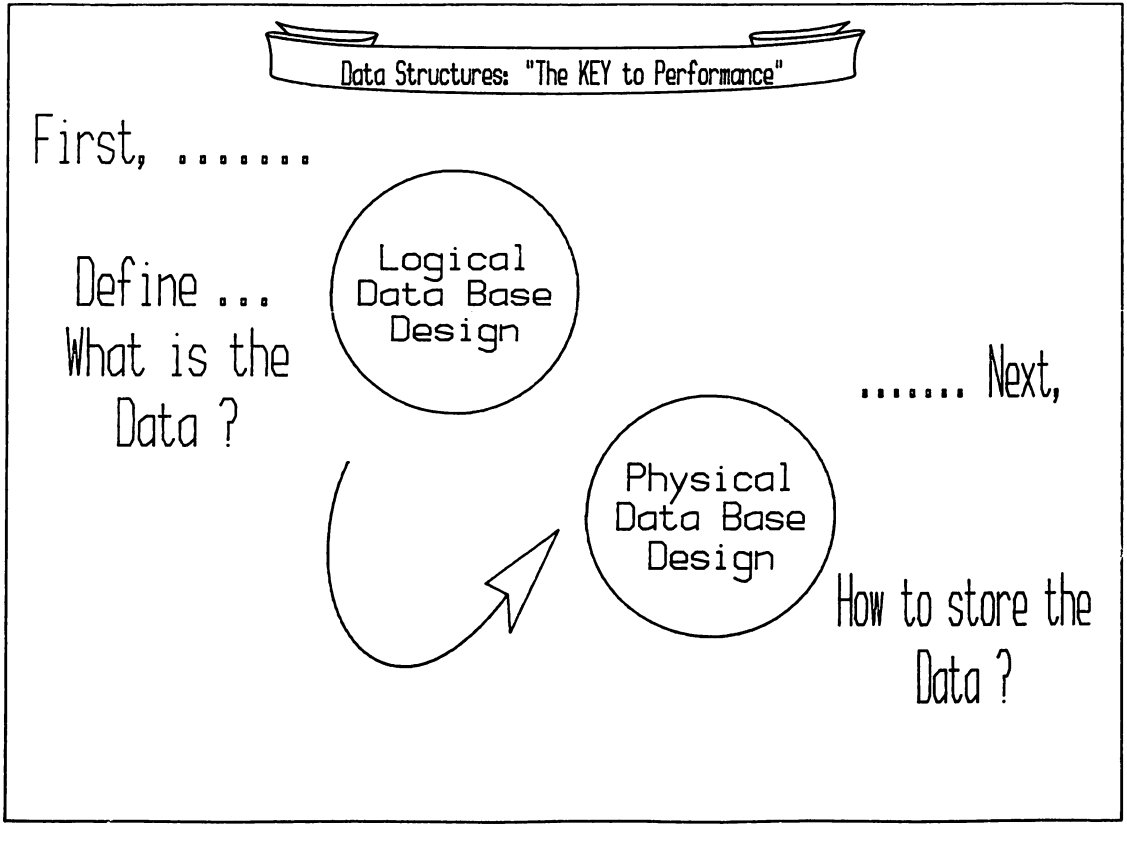
PowerHouse
Data Structures: "The Key to Performance"

There are many factors which play an integral part in the overall performance of an application. The objective of this tutorial is to identify why data structures are the "KEY" to efficient and well tuned applications. In order to do this design criteria will be introduced and discussed for optimizing your structures.

Some of the criteria to be presented are:

- Logical Data Base Design
  - Essential Systems Analysis
  - Data Modeling
  - Normalization
- Physical Data Base Design
  - Image DBMS vs Indexed Files
  - Blocksize & Capacites
  - Paths/Keys
  - Application Considerations
- Prototyping Application using 4GL PowerHouse

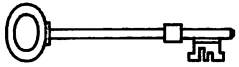
This design criteria introduced in this tutorial can be used for applications developed in other 3GL or 4GLs.


Data Structures: "The KEY to Performance"
First, ...
Define ...
What is the
Data ?
Logical
Data Base
Design
Physical
Data Base
Design
..... Next,
How to store the
Data ?

Data Structures: "The KEY to Performance"

Logical Data Base Design

Analysis Phase

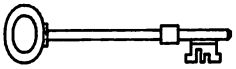


- Formal methodology
  - Essential Systems Analysis (ESA)
    [Top-Down Approach]
    - Data Flow Diagram (DFD)
    - Mini specifications
    - Data Dictionary
  - Blitz

  - Data Modeling
    - Entity relationships
      - Entity Relationship Diagrams (ERD)
        - 1 to 1
        - 1 to n

* 4GLs are perfect fit for Data Modeling

Data Structures: "The KEY to Performance"



Logical Data Base Design

Analysis Phase

- Informal methodology
  - Identifying data entities and their attributes
    - Normalizing data structures

      " Does each data element depend on its key, the whole key and nothing but the key ?"

    - Boyce-Codd Normal Form (BCNF)

      " Every determinant (attribute) must be a candidate key of the relation (entity) "

Data Structures: "The KEY to Performance"

CASE Tools - EXCELERATOR

. data analysis
. automates
. validates

LDBD

PDBD

PHLEX by Cognos

. PHDict Entities

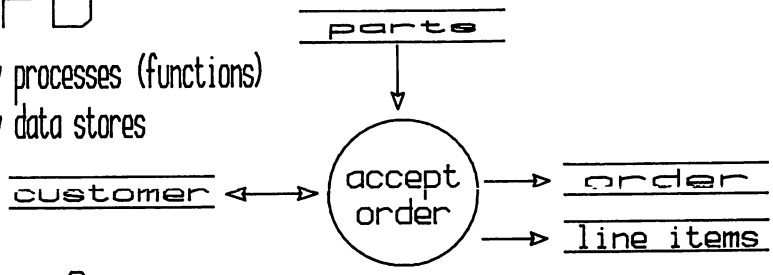
Link to

EXCELERATOR by Index Technology Corp.

Data Structures: "The KEY to Performance"

DFD

- Identify processes (functions)
- Identify data stores



Mini Specs

accept order

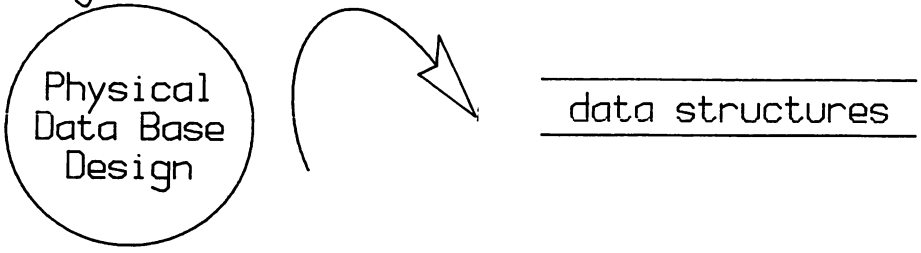
- For each line-item
  - multiple qty * price = extension

Data Dictionary

customer = [cust id, name, address, current balance]

Data Structures: "The KEY to Performance"

Design Phase

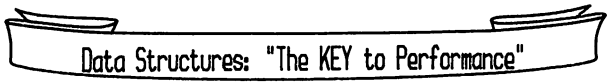


What data structures are supported ?

HP - MPE/Indexed/Image DBMS

DEC - Indexed/RMS/R db

DG - Infos



"Define data structures with PowerHouse in mind ..."

Design Criteria

- Data Base Structure
  - Image DBMS or Indexed (KSAM)
- Number of Paths/Keys
- Physical Data Structures
  - Blocking Factors
  - Capacities
- Application Considerations
- Other Software Alternatives

Data Structures: "The KEY to Performance"

" A database models the dynamic behavior of its entities and their attributes by means of entries"

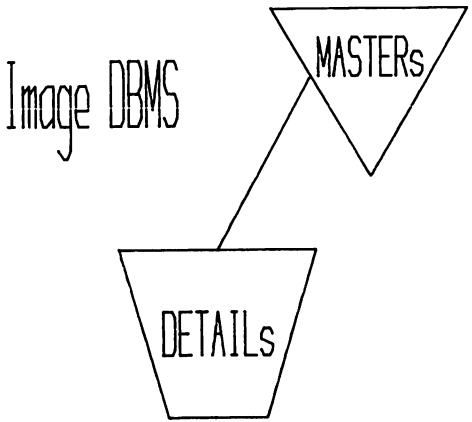
Alfredo Rego

- HP Image DBMS "Network Structure" with two levels of hierarchy
  - Masters
  - Details

What are the advantages of using Image ?
disadvantages ?

Data Structures: "The KEY to Performance"

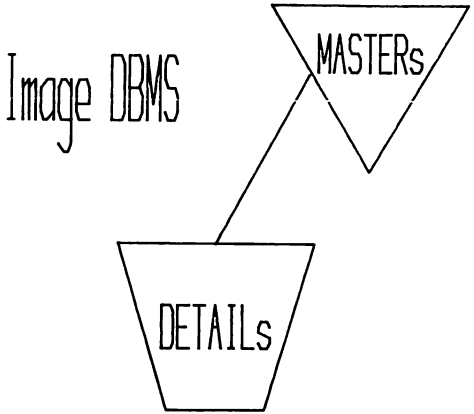
How can I access records in Image ?



- Unique KEY Masters
  - "Hashing" - contents-oriented access method
- Relationships for these masters stored in Details
  - "Chaining" - contents-oriented access method
- Image DBMS
  - Also supports address-oriented access methods
    - Serial
    - Directed

Data Structures: "The KEY to Performance"

Advantages of using Image DBMS as data structures !

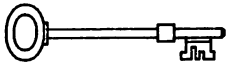


- Referential Data Integrity
- Backup & Recovery Features
  - Transaction logging
    - STARTLOG & STOPLOG for PowerHouse
- Image DBMS Utilities
- Third party Utilities
  - Adager/DBGeneral

Image DBMS

MASTERs

Data Structures: "The KEY to Performance"



Design Criteria for Masters !

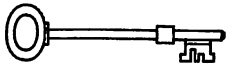
- Optimal Block Size (BLOCKMAX)
  - REBLOCK feature of ADAGER or DBGENERAL
- Set capacities to odd number
  - Improve "hashing algorithims"

- Keys
  - Types of X, U, and Z reduce migrating secondaries
  - For numeric fields < 5 digits declare as J1
    - PowerHouse type INTEGER size 2
  - * For numeric fields < 10 digits declare as J2
    - PowerHouse type INTEGER size 4
  - For numeric fields > 10 declare as PACKED size n

Image DBMS

Data Structures: "The KEY to Performance"

MASTERs



* J2 Master Keys

" ... If value does not exceed the capacity ..."

" ... Then record# of entry will equal to its KEY value ... "

Results

- Master set will contain no synonyms
- Disc space (sectors) will not be wasted
- Entries can be batch loaded faster than hash entries
- Physically order the entries in any desired sequence

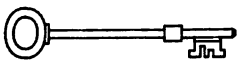
Mark Trasko,
"The Future of Database Technology"
Supergroup Association Magazine, July 1986

Image DBMS

DETAILs

Data Structures: "The KEY to Performance"

" Poor Master set performance ... affect Details"



Design Criteria for Details !

. Optimal Block Size (BLOCKMAX)

. Avoid more than two paths

. Avoid sorted paths if at all possible

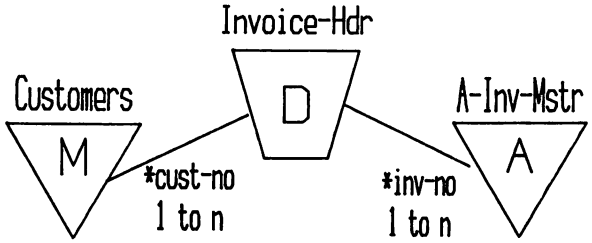
* . Determine optimal Primary Path

Image DBMS

DETAILs

" Primary Path .... increase/decrease retrieval access ... "

" Select most frequent PATH with more than one entry point "



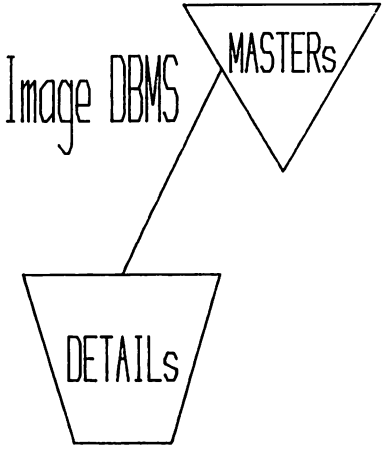
... Primary Path should be *cust-no in this example ...

Data Structures: "The KEY to Performance"

Disadvantages of using Image DBMS data structures !

" The Three Bears of Image"

Fred White



- Papa Bear - Integer Keys
- Mama Bear - Sorted Paths
- Baby Bear - No. of Paths

" ... use the above with considerable thought ..."

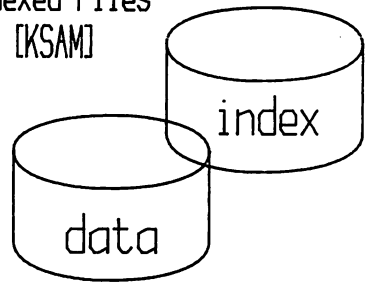
. Image DBMS allocates all necessary disc space "upfront"

. Without third party utilities can be time consuming for
  . file reorgs, etc ...

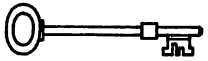
Data Structures: "The KEY to Performance"

How can I access records for Indexed Files !

Indexed Files
[KSAM]



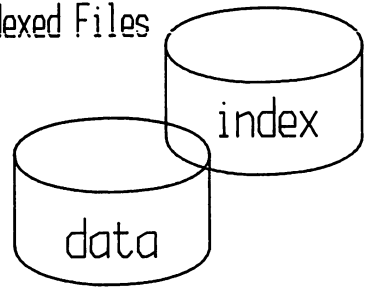
- KEY access
- Partial KEY (Generic KEY Retrieval) access
- Serial
- Directed



- PowerHouse products allow generic retrievals for QUIZ & QTP & QUICK

The good and bad of using Indexed data structures !

Indexed Files



" The good "

. more flexibilty
  . data structure easier to change
  . generic retrieval for character keys

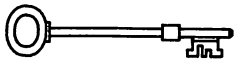
. less disc resources
  . data file allocated in extents

" The bad "

. increase memory requirements in multi-user environment

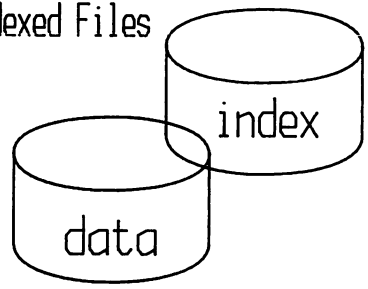
. no built in referential data integrity like Image Masters to Details

Data Structures: "The KEY to Performance"



Design Criteria for Indexed Data Structures !

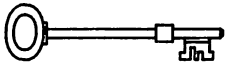
Indexed Files



" ... referential data integrity ... "

- Data integrity can be enhanced using PowerHouse
  - Master to Detail type relationships
    - Lookups
    - Linkage of screens
      - Passing/Receiving file
    - Append mode processing
      - Detail file

Application Considerations



- KEYS (Paths)
  - Number of keys
    - More keys increase inquiry capabilities
    - More keys also increase I/O in Updates
  - Concatenated keys (Indexed Files)
    - Retrieval of records in ascending order (key = [claim no + diary date])
- Security
  - May require additional data structures
    - QUICK Screen functional Menus
    - Password Security by ID
- Audits
  - Additional files to track (audit) changes, etc

Design Criteria for QUICK Screens !

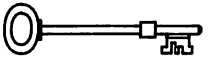
- Mirror data base design
  - For each Master a screen
  - For each Detail a screen
    - Same for multiple indexed files
      - Build in referential data integrity
- Exception is usage of Append Mode Processing
  - Detail Files
    - 1 to n relationship for two files on single screen
- Incorporate Locking Strategies
  - For multi-user environment
    - Screen LOCK Base/File or LOCK/UNLOCK Verbs
- Closing Files Explicitly

Data Structures: "The KEY to Performance"

. Develop/prototype application using Indexed structures
  . Easier to change structures

. Specify optimal blocking factor for data structures
  . MPEX - ALTFILE file-set,BLKFACT=BEST -

. Once data base design is accepted can change to Image

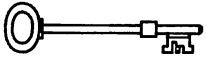


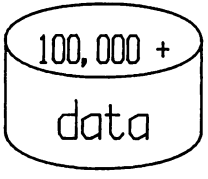
. Large Blocking Factors for batch processing
  . QUIZ and QTP serial reads

. Small Blocking Factors for screen processing
  . QUICK record retrieval

Other Software Alternatives

OMNIDEX / OMNIQUIZ by D.I.S.C



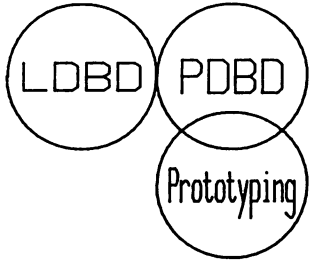


QUICK & QUIZ & QTP

- Advantage to end-user
  - Allows search on non-key fields, etc
- Advantage to data base administrator
  - Change search criterie without reorgs
    - Change only OMNIDEX structures

Data Structures: "The KEY to Performance"

" Prototyping Application"



Architect → Cognos, Inc.

LL'Spirit → Singapore Computer Systems

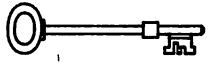
" Is feasible in a 4GL environment "

Data Structures: "The KEY to Performance"

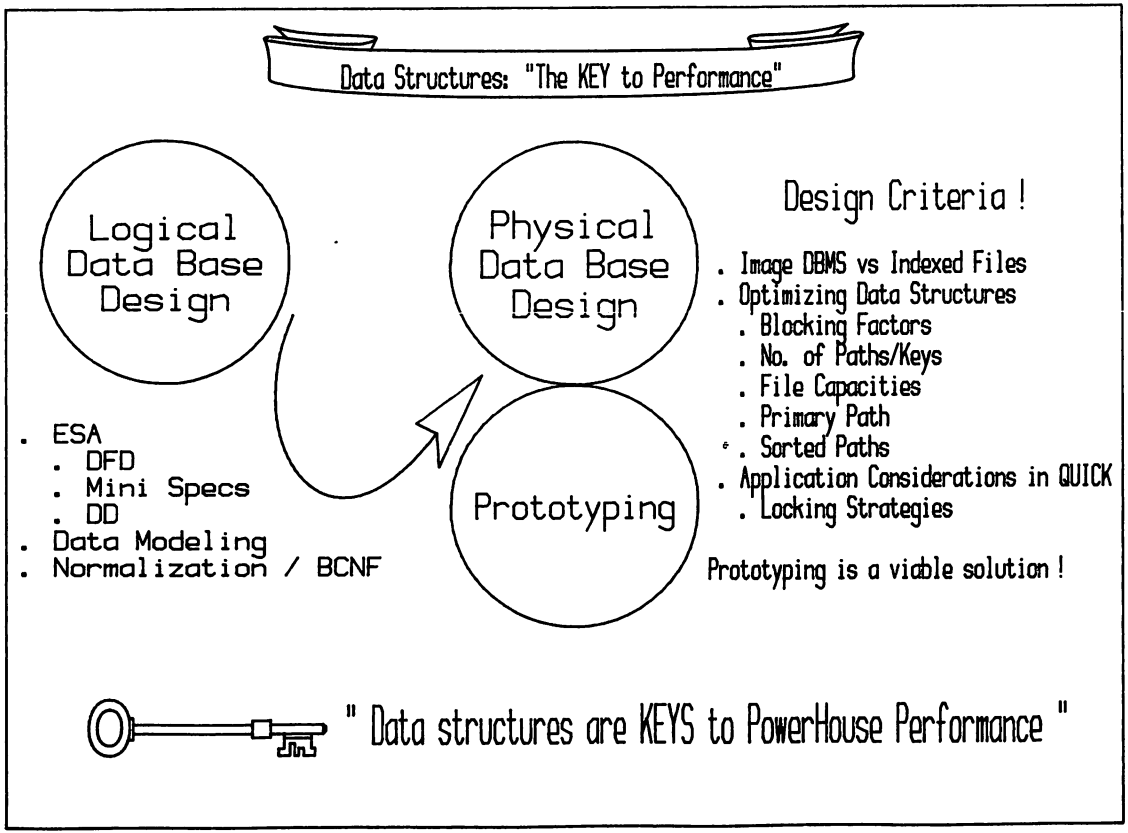
Advantages of Prototying in 4GL !

Prototyping

- Confirms users specifications (LDBD)
- Encourages users participation
  - validating inputs and outputs of the system
  - determine data base design is efficient or lack of
  - easier to evaluate then written specifications
- Shortens development cycle
  - Staff could be developing other modules at same time
- Prototype can be done on Micro computers using PowerHouse PC
  - After acceptance; code transportable to mini (HP, VAX, DG)



" committment from DP staff and end-users "


Data Structures: "The KEY to Performance"
Logical Data Base Design
Physical Data Base Design
Prototyping
. ESA
. DFD
. Mini Specs
. DD
. Data Modeling
. Normalization / BCNF
Design Criteria !
. Image DBMS vs Indexed Files
. Optimizing Data Structures
. Blocking Factors
. No. of Paths/Keys
. File Capacities
. Primary Path
. Sorted Paths
. Application Considerations in QUICK
. Locking Strategies
Prototyping is a viable solution !
" Data structures are KEYS to PowerHouse Performance "

Data Structures: "The KEY to Performance"

Author:

David G. Robinson is considered one of the leading authorities on PowerHouse software. He is founder and general partner of PowerSpec International the world's premier PowerHouse training center. David is an active member in the user community having presented many techical papers on PowerHouse in Europe and the United States.

He is also co-editor of TNT, a quartlerly publication of Tips, News, and Techniques on PowerHouse, which is distributed world wide.