

# **Data Expert System**

## **An expert system for the database with idea of new generation language.**

Pramila Dangwal

Institute of Technology Management,  
Defence Research & Development Organization,  
Mussoorie, Uttaranchal.  
pramila\_dangwal@rediffmail.com

**Abstract:** Expert System has a formal definition of system having collection of data from the expertise of particular field on specific topic. This research work is on the expert system of data. Here expert means a very peculiar system to manipulate & maintain data. We are well aware of the word DBMS & RDBMS, which stores the data in two dimensional forms but this expert system works on the collection of multidimensional data having qualities of data warehousing, mining in a grid form of allocation. This feature of the system helps to enhance the relation among data stored in a non continuous (grid) form. This non continuous allocation reduce the replication of data in system. Therefore it also works for well utilization of memory. This system uses 5GL features for data entries and manipulation which support system by less time consumption in above mentioned process, in different languages. 5GL features are used to improve user system utilization by giving the feature of Library and Micro construction. For this purpose the research work has an idea of merging 4GL with 5GL. That is this feature of the research work presents the paper work on the idea of implementing PL/SQL by 5GL. Therefore it will reduce the task of user to learn or permutate with 4GL in database. This work also has some attention to DL & DRO to support the objective of system & its functioning.

**Keywords:** DBMS: Database Managemen System, RDBMS : Relational Database Management System, 4 GL : 4<sup>th</sup> Generation Language (SQL, PL/SQL), 5 GL : 5<sup>th</sup> Generation Language (Speech-data conversion language), DL : Digital Libraries, DRO : Digital Resource Organisation.

### **1 Introduction**

An Expert Data System is a paper work on the Database System to make it work as an Expert System with the implementation of expert ideas of different systems to collectively enhance the performance of database system.

Therefore it is a database system practiced by the expert advice of expertise in different system to enhance their systems performance. This collective work of ideas itself forms an expert system that is termed as Data Expert System.  
(Note: To explain this idea we use here a grid form of explanation (i.e. three dimensional) & will be explained in a form of phases say phase1, phase 2 and so on.)

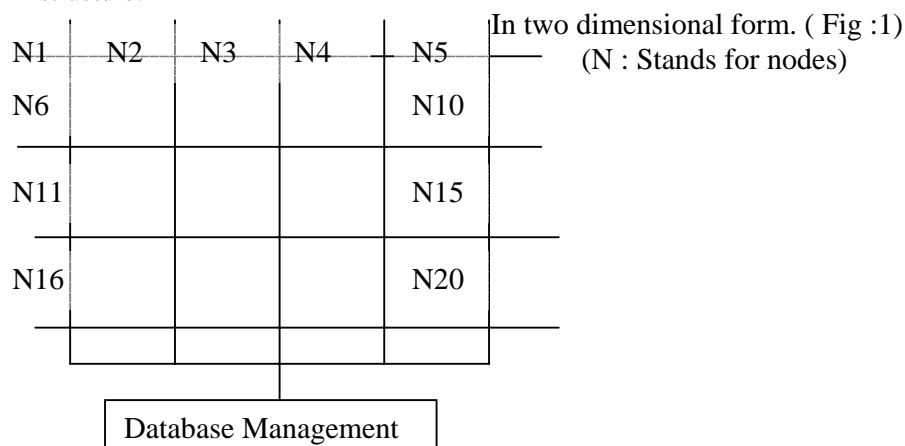
### Phase 1:

#### **This phase consist of storing process of data**

As we all know till now we have been practicing on two dimensional data base allocation i.e. table form of structure and to maintain a relation among them there was a lot of memory allocation by replication of data. Therefore this expert system has an idea of allocating of data in multidimensional way i.e. a grid form of way resulting data is stored in a grid kind of structure.

**Storage** : This is more or less having storage concept as used in Oracle 10g( g stands for grid) or a power of grid features . A power griding(grid) is a technique used for the well utilization /distribution of power among different channels . Each and every require terminal is attached to one of the multiplexing joint of grid & power of any capacity can be absorbed by any terminal form the system , at any time , without any boundary conditions. Here the grid structure provides some more or less same kind of feature. Inspite of having different tables in order to maintain a good relationship among related table in a database. This database will have a table that will work or function as a database in itself. That is it will help to have a function as a database in itself and will help to have the features of database in a table.

The multidimensional / grid form of allocation of data or in other words in tables with the grid structure each node in a grid will work as a column in a table to store a particular kind of data and row of each entity will not have a specified structure but an ordered field have related attributes orders & discipline. This ordered field attributes for different entity gives an idea of two dimensional structure of database but this ordered field will be for record purpose to maintain a discipline while processing , retrieving or surfing for information from the table. The table will have the following kind of structure:



In three dimensional form (Fig2) ( $N_n$ - Node in data structure for field in table)

Single cube allocation in memory

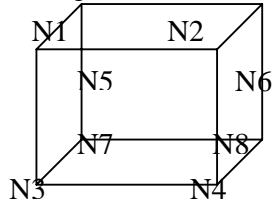


Fig :2.1

Two cube allocation in memory

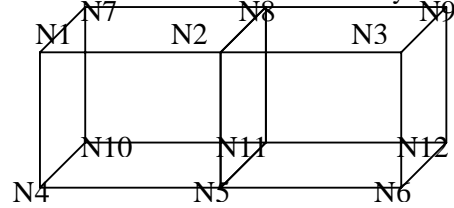


Fig 2.2

Three cube allocation in memory

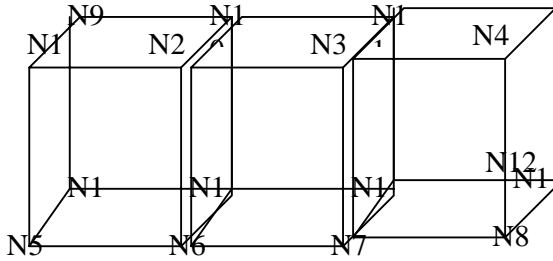


Fig :2.3

Four cube allocation in memory

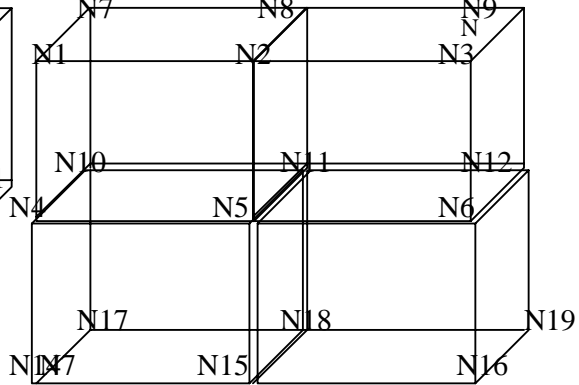


Fig 2.4

Actual form of multi cube allocation in memory in multidimensional way

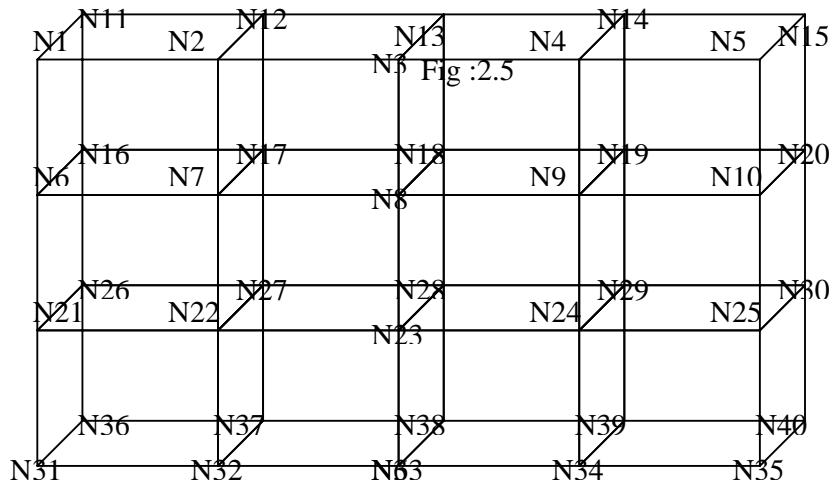


Fig :2.5

So it is hybrid /grid kind of structure for the field allocation of table in application .As the allocation node sequence is considered it is more or less in the form of multidimensional array so the input can be taken by the multidimensional data structure source.

But when it comes the case of memory allocation a special application ( with very brief strategy) can be used . The application will work on strategy described as follows:

If the input values are 8(i.e. fields) or less then eight , one cube numbering nodes as in figure 2.1. So first, it will check the multiplicity of number of nodes allocation with eight and also check whether it is less or more. If the number of nodes are multiple of eight then the quotient formed by divisor (i.e. eight) & dividend (node) will be the number of cubes allocation. If the node number is not a multiple of eight then a one added cube to quotient is taken to accommodate the remainder nodes. And allocation of nodes as described in figure 2 (1), (2), (3) & (4) will advantage of reducing complexities from table.

Second is the arrangement of the cells. Cells will be application/programmed to arrange the square structure of cube if possible and if not possible then the priority will be given to row then column. That is columns will be arranged in memory in a form of matrix in which row is given priority then the columns for example if nodes are eight then cube allocation will be one , with nodes allocation as mentioned in figure 2.1.

Therefore the strategy used for the allocation of nodes is as follows:

The number of cubes (always even)	Number of nodes (n=0 to inf)	
1) 1	$0 < x \leq 8$	4
2) 2	$0 < x \leq 12$	6
3) 4	$12 < x \leq 18$	6
4) 6	$18 < x \leq 24$	6
5) 8	$24 < x \leq 30$	6
6) 10	$30 < x \leq 36$	6
2n	$m < x \leq m+6$	6

Therefore the strategy for cubes allocation with the allocation of nodes follows the above. These cubes combinations offer multiple paths for the different fields of a table to process its data to the database management system. Therefore it will help in enhancing the processing speed of queries.

Therefore this feature of hybrid storage of data in memory works on improving the speed of query processing due to readymade relationship

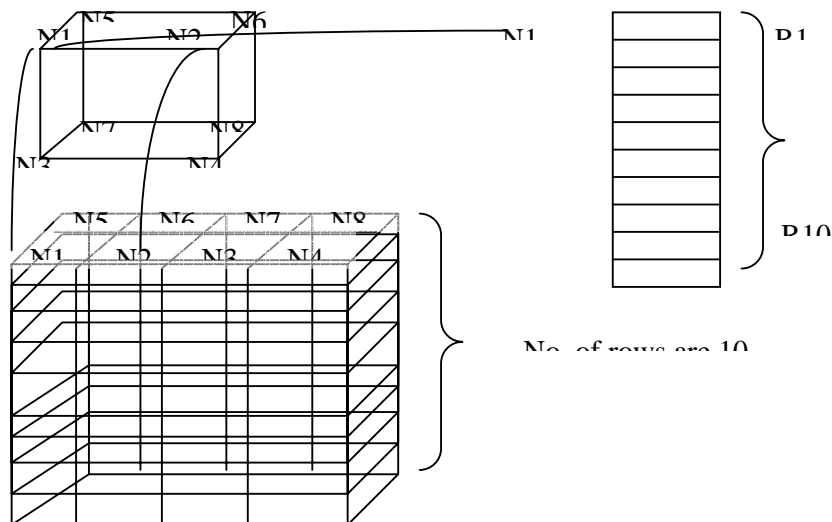
among fields of a table & by multi transmission way. Easy programming in PL/SQL & micro in related fields.

So here the data allocation is more or less like data warehousing dimensional model, data cubes. Therefore it offers all the advantages (almost all) of data warehousing. This storage patterns covers data warehousing completely. Therefore this system can be called as a system having the advantages of database management system & data warehousing.

Sliced operation of data.

The associated figure is as follows:

$N_n$  : Fields of a table(say table has 10 records)  
 $R$  : No. of rows in the table (let be 10)



As it supports all features of data warehousing features therefore it can support all of its tools like OLAP(Online Application Programming) . Therefore it will that is its this structural allocation feature will support the feature of data warehousing like:

☞ **Data Extraction** : From variety of stored data like metadata, production data etc that are stored in single table here.

☞ **Data Cleaning** : Its quick and tricky like by transformation rule, domain specific knowledge, performing parsing & fuzzy matching auditing etc.

☞ **Data Transformation** : To transform data from one allocation to another.

☞ **Loading** : In batch, sequential or incremental etc form.

And like warehousing storage can be used / programmed by OLAP (Online Application Programming) that will accept command from user simply in oral form etc.

These structures offer following kinds of advantages in database application

1. It automatically offers relationship between almost all nodes of the tables. Therefore without creating relationship among nodes we can access the advantages of relationship by only specifying some integrity rules with respect to fields.
2. Reduces the complexity of surfing all tables with common data for specific data in database that is this hybrid structure can accommodate all the tables data having common field. So this structure of allocation will help in applications for having less complex program for surfing purpose.
3. Macros & Library functions can be created on the associated nodes which due to the interrelation will be less complex, hazard and of reduced size.
4. As it has a structure of multidimensional on the solution methodology of multidimensional matrix can be also used on such table construction.

*Advantages over memory allocation:*

1. Ordered structure so the data can be orderly arranged in memory to avoid the confusion in static as well as dynamic allocation in application.
2. It can be allocated in memory by multidimensional array by any application (PL/SQL(4 GL) , High Level Language(3 GL) or by (5GL).
3. It will avoid the replication of data in database in comparison to the existing one due to repetition of common field in different tables.

*Advantages over processing:*

1. It will require less time to surf the required data.
2. It will require less time in processing it's queries because of interconnectivity of each node with each other that offers multi path for a data to process through the channels.
3. Memory allocation advantages offer less time consumption for application to retrieve data from memory

*Disadvantages:*

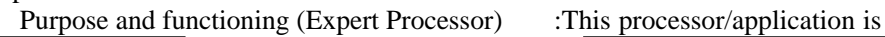
1. For back up every time a fresh table would be required & this is more time consuming. Only a backup of some portion of table won't satisfy the table and its requirement as the relationship among various field won't be well backup if only a portion of it is backup. In all only a portion backup will not satisfy the relationship strength further.
2. Recovery of data is slightly complex then earlier one.

But these two disadvantages can overcome by the application strategy of OOPS(Object oriented language), PL/SQL etc further if required. Due to its this storage procedure it will work as perfect storage allocation in Knowledge Management & serve all of its task up to the level. The above mentioned was an allocation concept. Next is work on processing concept covered in Phase II.

### **Phase 2: (This phase is concerned with the processing of queries)**

**Processing** :In DBMS the allocation solution only cannot result to a good & effective DBMS it requires an equal support on an effective processing system. A DBMS can be called as RDBMS or a knowledge System or an Expert System it has an effective method of query processing. These both i.e. storage & processing although are stand alone concepts but in order to form an effective database system they work collectively to give a desired result. Till now we used to use 4GL (i.e. Query processing language) to deal with stored data. And DBMS or RDBMS are now a days joined/merged with compiler using 4 GL language to manipulate & mine with data stored. In this 4GL (Generation Language) we have some key words that have some specific task in the package (i.e. key words are ready made functions that can be called for the completion of some specific task). When these key words are arranged in a specific order a specific task is done by Database Management Application System. Like displaying, filtering, printing, creating, viewing, updating, surfing data's with some clauses from database. This technique of communicating with database is called DDL (Data Definition Language) or Query Processing Language or SQL(Structured query language). Next is to create user own functions that can manipulate with the data of database. Earlier we talked about key words that are the ready made functions for database management system. But this management system is also offered with the facility of creating functions that works according to users wish. The programming language in which it is written is called PL/SQL (procedural language/structured query language). This feature of DBMS is called DMS (Data Manipulation Language). So in DBMS we are offered two ways to communicate with stored data one is DLL and second is DML. For this purpose we require a query processor that works on converting 4GL language to machine understandable language. That is here query processor act as a translator or communication mode between the user & data storage location or DBMS. But as we all are aware of this fact that 5GL is been introduced that offers that facility to its clients, to enter words in text spoken by them. Here we merge 5GL with 4GL to rise the usability, efficiency, validity, user friendliness & convenience of database.

The strategy of linking of 5 GL translators to 4 GL translators i.e. query processor and this query processor to database management system as follows:

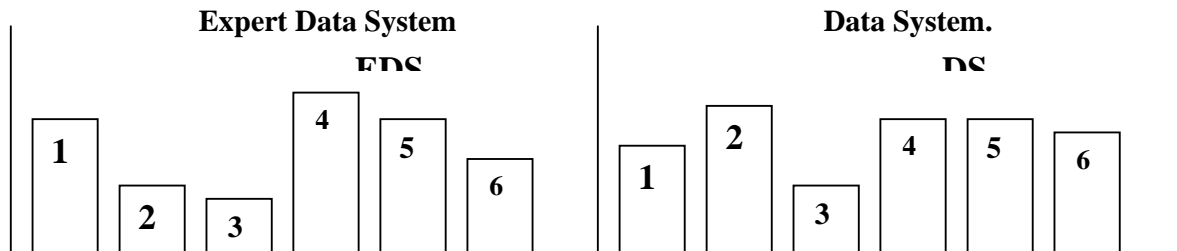


## QUERY PROCESSOR

```
graph LR
    User[User] -- Audio --> IO[Input/Output]
    IO <--> 5GL[5GL Translator/processor]
    IO --> D1{ }
    D1 -- YES --> EP[Expert Processor]
    D1 -- NO --> D2{ }
    D2 -- YES --> DBMS[(DBMS)]
    D2 -- NO --> IO
    DBMS <--> QP[Query Processor]
    QP --> IO
```



Therefore it is the flow chart of process flow in the system:



## Phase2.1

Statistical comparison between both systems:

Here in this statistics:

1. Time required for the completion of task.
2. Efficiency required to learn the language.
3. Input Consistency.
4. Processing Task.
5. Validity( valid to user requirement and simplicity)

Complexity of running the process.

**Expert System:** This expert system is an application that can be designed in any 3GL language to convert 5GL to 4GL . Here some mathematical concept will be also used in order to complete task. As we all know the application can be constructed in many ways as per user talent to maximize the task with minimum (least) strategy. Here is one of the solution for this problem:

When we enter a specific line in the 5GL mode and it is transmitted to 5 GL transmitter which convert the audio message into text form and then display it on screen. After the user satisfaction with the syntax of data entered by it in an audio form & it is displayed on output device, it gives the signal to pass the displayed syntax to Expert Application. Once the signal is processed the displayed (current) text is sent to expert system for further manipulations or in language of application construction on giving an enter signal on expert procedure call is made. This enter call or procedure call passes the displayed text as input on which procedure will do operations. So by enter signal 5GL text is passed as value to the expert procedure. This value is stored into a string array of undefined size. Now every time a call for expert procedure is made a following operation is done in the application:

It takes one more array of undefined size. Next, array two will enter the text data that exist in array one (original message) word by word taking space as a recognizer of word in syntax until and unless all words of the original message, are passed to array two. That is a loop will be there in this portion of application that will run until and unless the receiving array i.e. array one in empty/blank.. In this way we will separate words from the syntax. Now every time a new word is formed it is again added by a loop in the series of array2 till it has all the words from storage location of its match(for this purpose there will be a (create) library in which we will give alternate SQL word for each word in English/normal language. If required in English then no alternate is to be used as this library will be having respective SQL words for English words but if the syntax is in some other language then again another translator in that specific language is required as to convert the syntax of given specific language into English).Now when application collects all the alternatives in SQL for the words in original syntax i.e when the alternatives in SQL for all words in array1 is received these words are stored one by one in new array say array3.Again it is acted by a loop for the possible probable syntax of array3 word. A most appropriate syntax is then again passed to the output for the purpose of displaying it on screen. Once SQL statement is processed on again by enter or after the user acceptance of its SQL statement (by enter)a call for query processor is made to work on the passed query for further operation. Here we can add another procedure which is exceptional that is on the choice of user to use it. This procedure on call will directly transfer the most probable SQL syntax from expert system to query processor by the command double enter after receiving the syntax from SQL translator directly. This procedure will save the time of the user but the accuracy of the result will not be as sharp as the earlier one that has one more step to go instead. So this procedure is exceptional and its construction strategy is not covered in this paper send to you. After the query is been processed to the query processor it will work as it does in ordinary database system.

## **Phase 2.2:**

**Process:** Application (expert processor) will receive a text in an undefined string array say A1. After, array A1 will be acted by a loop till the number of times it is equal to the number of words available with array A1 and these words will be one by one stored in another undefined string array say A2 .When array A2 receives the number of words equal to the number of words in array A1 it will be again acted by another loop No sooner A2 receives a blank form an array one (A1) it will call for an another procedure which will enter an alternate meaning for string A2 from library to another undefined string array say A3.This will be available SQL word for the given text word. This procedure will store the available SQL word for the given text word.

This procedure will store the available SQL words in no. of undefined array. This recursive process is processed until and unless the array A1 is empty and A3 have a number of permuted array respective to array A2. These SQL words of array3 (A3) are now acted by another procedure say permute that will have all the possible permutation of the words of array (A3). These array formed by the permutation are again added by a procedure that checks for the best SQL syntax among all & once best SQL syntax is formed it is displayed on screen ( here for this purpose we will have another library in which we will have all the possible SQL syntax with which these permuted syntax is matched for the accuracy) . Once displayed on screen and received external command this SQL syntax is sent to Query Processor for further operation on request. As whole process is computerized and time required by huge processing task is slightly more( which is key factor in today's environment) but keeping in mind the emerging processing speed of the computer, multi way processing capacity as described in phase1 and its advancement in user friendliness of system( as there will be no need of learning DLL, DML or procedural language) make it to over come the defect of time factor. It will serve us an easy way to communicate with our database and construct queries, micros, and library packages in DL & DRO.

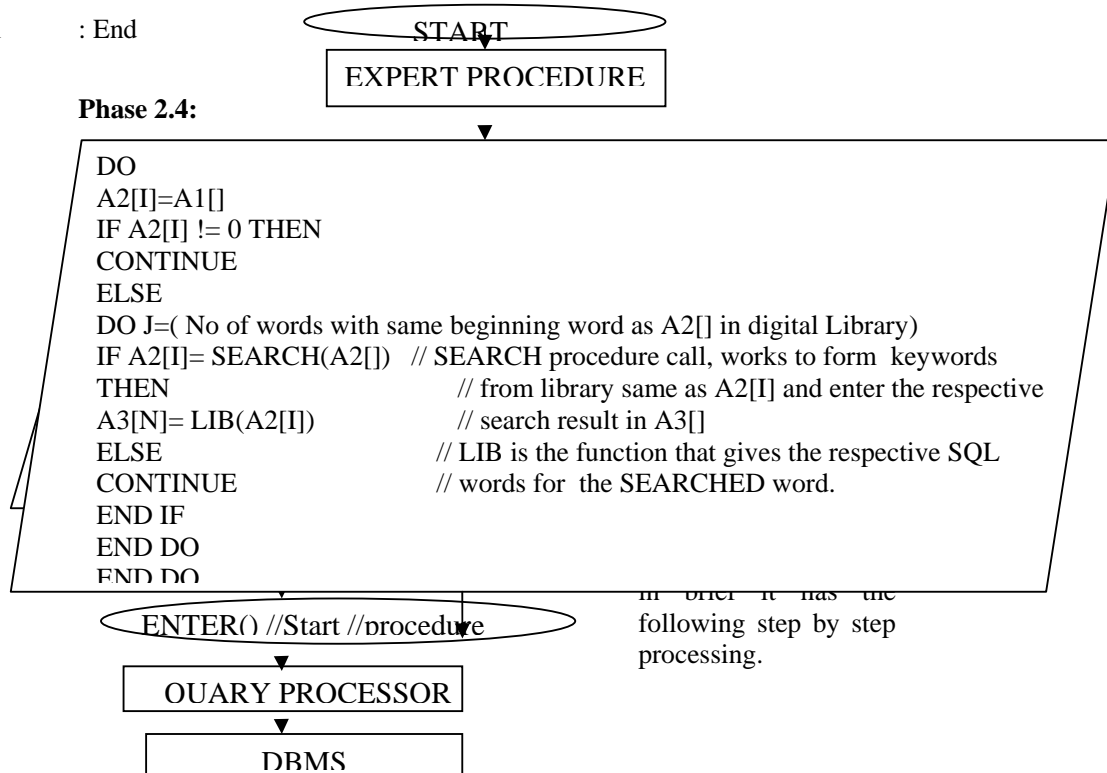
### **Phase 2.3**

#### **For Expert System:**

- Step 1 : Input is taken from the input sources in the audio form that is transferred to the 5 GL translator for the respective text conversion of it which on again receiving input(enter)(application of expert system receives input from here) is stored in a string array of undefined size say A1.
- Step 2 : This text is processed by a loop to determine the no of words in an array A1 say I.
- Step 3 : Run a loop for I number of times.
- Step4 : Store the first word of string array A1 in another undefined sized string array A2.
- Step 5 : Call for the search function . This function performs the task of searching the text word matching to array A2 form the dynamic library that has the ciated SQL key words to the text words.
- Step 6 : Use if conditional function for the matching text key word formed by search function form DL. If the word matches with the array A2 then store the associated SQL word in new array A3 by the library procedural call.  
( Library procedure performs the function of carrying the respective SQL words from DL that matches with string A2.)
- Step 7 : End loop.
- Step 8 : Call for the procedure permute.  
( Permute is a procedure that checks for most appropriate combination of array A3(SQL words).

Note : This is itself a new application that requires a mathematical strategy for its solution & the strategy is permutation that surfs for the most appropriate SQL statement among permuted ones.

- Step9 : Most appropriate (SQL)/A3 combination is after sanded to the output device. ( Here the help of another DL is taken which has database of all possible queries in database. This procedure matches all possibilities of combination of A3 with all existing SQL queries in database & select the most appropriate combination of A3 as final query. As the process is lengthy and the matching of every data with database elements of library is done, the process strategy is slightly time consuming . But for this purpose some more techniques can be used to reduce the time factor in this procedure).
- Step10 : On receiving enter signal from input signal from input device this final query is send to the query processor for further processing.
- Step 11 : End



**Conclusion :** Therefore it is a brief representation of my research work on database management system to make it Expert Data System . This database as per expert system has an expert ideas of different researchers working in Oracle. Microsoft & working in database enhancement concept. It has

different ideas from experts like grid structure, multidimensional allocation, retrieving & processing techniques etc. These expert ideas are used in different arrangement here in this research work to enhance the performance of DBMS as it is used to do in formation of new advanced version of DBMS. This advancement in DBMS will help in better memory utilization & better processing speed of an application.( this is covered in first phase). The second phase work on the concept of increasing the validity, user friendliness & reliability of application (DBMS). This phase uses the combination of 4GL & 5GL with additional application naming Expert Processor which is working as an interpreter of converting 5GL to 4GL. So here in phase two an idea of expert processor is introduced that has above mentioned goal with algoed strategy. This phase has four sub phases that consist the location of system (expert system), functioning of system, and flow of data in system and strategy/algorithm used in its construction.

So resulting these both phases collectively can give a new version of DBMS/ RDBMS which can increase the use of DBMS and can be new with better replacement to the existing systems giving the up to the mark satisfaction of expert system of database well called as Data Expert System.

#### **References :**

- D Marco(2000), *Building & Managing the Metadata Repository* , WILEY  
D Meyer(1998), *Building a Better Data Warehouse* ,PRENTICE-HALL  
W Inmon(2002),*Building the Data Warehouse 3rd Edition* , WILEY  
W Inmon(1999), *Building the Operational Data Store* ,2nd Edition ,WILEY  
E Vitt(2002), *Business Intelligence Strategies*, MICROSOFT PRESS  
M Whitehorn(1999), *Business Intelligence: The IBM Solution - DataWarehousing and OLAP*, SPRINGER-VERLAG LTD  
D Pyle(2003), *Business Modeling & Data Mining*, MORGAN KAUFMANN  
WN Venables, BD Ripley(2000), *Statistic & Computing(SProgramming)*, Springer  
Geoff Sharman(1987), *Introduction to Database on Micro Computer*  
Emerson & Darnovsky(1984), *Database IBM PC*  
Paul Grefen, Babara Pernici & Gabrid Sanchez(1999), *Database Support for Workflow Management*, KAP  
Prakash (1990), *Data Management(Current Trends)*,Tata Mc Graw Hill  
Alan Mayne(1983), *Introduction to Relational Database*, Michel B Wood  
James Martin (1985), *Principals of Database Management System*, Prentice Hall  
Loomis(1990), *Database Management& File Structure*, II edition, Prentice Hall

Hansen & Hansen(2000), Database Management System, II edition, Prentice Hall  
 Richard Twatson(1996), *Datamanagement in Organisational Perspective*, Prentice Hall  
 PH Martin(2000), *Computer Database Organisation*, II edition, Prentice Hall  
 Loomi(1983), *Data Management & File Processing*, Prentice Hall  
 S Ramani, R Chandrasekar(1987), *Knowledge Based Computer System* , KS Publications, Anjanequler NHP  
 Louis Ermine1995,,*Expert System, Theory & Practice*, Prentice Hall  
 Jean James Bradly(1990), *An Elementary Introduction to Database Management System*, Dryden Publications  
 Val Occarde(1993), *Relational Database Theory & Practices*, BPP  
 JS Edwards(1992), *Building Knowledge Based System*, Wheeler Publication  
 PN Rastogi, Knowledge Engg & Expert System, BPP.  
 Andrews S Thane(1985), *Structured Computer Organisation* ,Nbaum  
*Computer and Information Processing(1983)*, II edition Wiley  
*Going Digital(1997)*, The Economist  
 Turban, Mc Lean, Wetherbe, *Information Technology Fro Management*, Second Edition, Wiley

--- ---- ---