A Mapping and Ranking of Selected Database Application to DBMS Models

Sili Li
Ghasem S. Alijani
and
Kai Koong

Southern University at New Orleans
Graduate Studies in Computer Information Systems
6400 Press Drive
New Orleans, LA, 70126

Abstract

Database Management Systems (DBMS) have changed significantly during the last four decades. There are currently six different database models: file processing systems, hierarchical, network, relational, object-oriented, and the object-relational DBMS. Even though all these models share common features, each model can be identified for its unique characteristics. Furthermore, implementation of the common and the unique features can produce a wide variety of database applications. Such a variety calls for a systematic approach for evaluating and selecting a database application. Based upon the six database models, a total of 33 features were identified from the literature. The findings indicate that the database features and the attributes of 28 popular database applications could be mapped into a schematic model and a three-tier logical structure accordingly. One the strengths of this study is inherent in methodology applied, because the technique uses a robust approach for identifying, mapping, and ranking decisions that is replicable. Database vendors, buyers, project managers, systems analysts, software developers, data administrators, and research scholars specializing in data modeling and applications should find this study useful.

Keywords: Schematic model, three-tier structure, mapping and ranking database applications.

Database technology has grown substantially in almost all application areas. From the small entrepreneur to multinational enterprises, some form of database applications is being used to support daily operations. According to a 1999 survey reported by the Gartner Group (2000), the database management systems (DBMS) market has grown to almost $6 billion.

Database management systems were first introduced during the 1960s and have evolved into six major models since. Generally, a DBMS model can be implemented using a file processing approach or a database approach. The file processing approach is defined as a collection of application programs that performs services for end-users such as the production of reports (Connolly and Begg 1999). Each file defines and manages its own data and it is impossible to associate data items or establish relationships among the files.

The database approach consists of the hierarchical (Burleson 1999), network (Larson 1995), relational (Yazici and George 1999), object-oriented (Tao, Y. and Grosky, 2000; Chung and et al. 2001; Vandenbussche and Paredaens 1995), and the object-relational data models (Muller 1999; Grimes 1998). In this approach, a single repository of data is maintained. It is defined only once and can be accessed by various users (Elmasri and Navathe 2000). It has been estimated that the development time using a database approach can be 4 to 6 times faster than that of a flat file system. According to Teorey (1999), the motivation for using databases rather than files is that they provide a variety of advantages such as the integrating of data and allow the processing of complex transactions with minimum redundancy.

Problem and Objective Statements

Presently, there are six different database models. These models are classified as file processing systems, hierarchical, network, relational, object-oriented, and object-relational DBMS. Each of these DBMS supports a unique set of applications even though they share common features. Selecting a proper DBMS and database application (products) can be a difficult decision because each of the features also provides a variety of options that may not exist in other products.

Some of the DBMS features are also used to develop new database model (Apicella 1998; El-Rewini and Hamilton 1995). Therefore, a mechanism is needed to assess the features of each DBMS model that are critical to the development of future models. Moreover, database applications should also be evaluated to assure that all the features of the DBMS have been implemented as required by the operating environment. Such a methodology would be beneficial to developers of new database structures. For users, the availability of a structured evaluation methodology can provide a guideline for selecting an appropriate product.

This study uses the existing frameworks of six DBMS models to map and rank popular database applications that currently available in the market. The findings of this study indicate that the products could be mapped to a schematic model and a three-tier ranking structure. The result of this study should be useful to database vendors, buyers, project managers, systems analysts, software developers, data administrators, and research scholars specializing in data modeling and applications.

2. METHODOLOGY

Clearly, DBMS have changed dramatically in the last four decades. Each succeeding DBMS model includes more features than the previous generation. However, there is a trade-off between efficiency and complexity when an attempt is made to develop the next generation of DBMS. Furthermore, user application requirements should always remain a major factor in the design and implementation of a database model.

The methodology used in this study consisted of a robust approach for identifying, mapping, and ranking decisions that is replicable. First, the schematic model is derived from a review of 33 features that are used for developing database models. Second, the mapping of the features provided a systematic approach for assessing and ranking the attributes of 28 popular database applications.

Figure 1 depicts a process for the development of a three-level schematic model for classifying DBMS features. This schematic model constitutes the initial framework used in the second part of this research that is focus on mapping and ranking the database applications. The most important procedure here is to identify the features and to classify them into levels. The classification is conducted by examining the percentages of DBMS that provide the specific feature. The most common features among the database models were classified as primary. The least common features contained in each database model were classified as unique whereas the remaining were classified as secondary. The following abbreviations are used to represent the respective databases: F.P. (File Processing System), H (Hierarchical DBMS), N (Network DBMS), R (Relational DBMS), (Object-Relational DBMS), and OO (Object-Oriented DBMS).

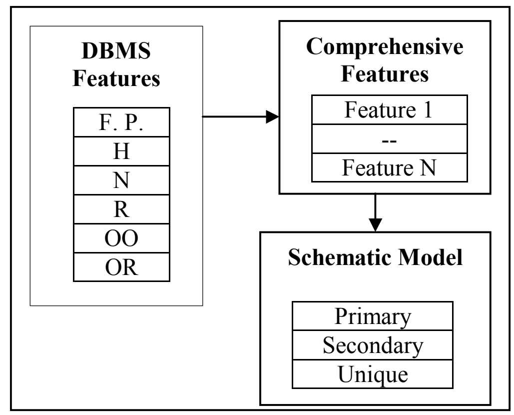

Figure 1: Process of Development a Schematic Model

Figure 2 shows the process for evaluating the database applications using the features that were identified in the schematic model. This process involves the actual mapping and ranking of the features of the selected products. First, the identified features in the respective applications were classified into primary, secondary, and unique features. This classification was conducted by examining the percentages of features in each of the products. The most common features among all the products were classified as primary and the east common were classified as unique. The remaining features were classified as secondary. The results for each product were then tallied and ranked.

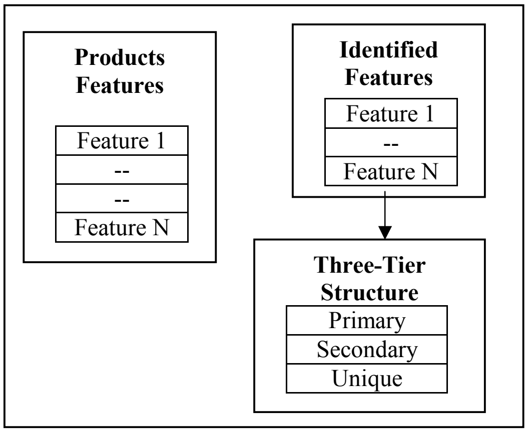

Figure 2: Process of DBMS Products Evaluation

3. FINDINGS

The outcomes of four major processes reported in this section include:

1) Identification and classification of the DBMS features.
2) Evaluation of selected popular database applications.
3) Mapping of the products into the schematic model.
4) Ranking of the applications by percentage of available features.

Table 1

Distribution of Features Among the DBMS

No.	Features\DBMS	F.P.	H	N	R	OR	OO	Total	Percentage
1	Storing, updating, deleting data	1	1	1	1	1	1	6	100
2	Availability		1	1	1	1	1	5	83.33
3	Controlling redundancy		1	1	1	1	1	5	83.33
4	Data independence		1	1	1	1	1	5	83.33
5	Economies of scale		1	1	1	1	1	5	83.33
6	Enforcing integrity constraints		1	1	1	1	1	5	83.33
7	Enforcing standards		1	1	1	1	1	5	83.33
8	Establish relationships		1	1	1	1	1	5	83.33
9	Existence of a catalog		1	1	1	1	1	5	83.33
10	Flexibility		1	1	1	1	1	5	83.33
11	Multiple user access		1	1	1	1	1	5	83.33
12	Recovery and Backup		1	1	1	1	1	5	83.33
13	Security		1	1	1	1	1	5	83.33
14	Complex data relationships			1	1	1	1	4	66.67
15	Easy data access				1	1	1	3	50.00
16	Manageability				1	1	1	3	50.00
17	Very large databases				1	1	1	3	50.00
18	Data persistent					1	1	2	33.33
19	Distribution					1	1	2	33.33
20	Reliability					1	1	2	33.33
21	Scalability					1	1	2	33.33
22	Supporting complex objects					1	1	2	33.33

23	Data structure representation						1	1	16.67
24	Encapsulation						1	1	16.67
25	Extensibility						1	1	16.67
26	Information hiding						1	1	16.67
27	Inheritance						1	1	16.67
28	Low impact of a failure	1						1	16.67
29	Polymorphism						1	1	16.67
30	Rule system						1	1	16.67
31	Simplicity	1						1	16.67
32	Small size	1						1	16.67
33	Versioning						1	1	16.67
	Percentage	12.12	39.39	42.42	51.52	66.67	90.91	100	

Table 1 shows the DBMS features that are used to derive the schematic model. This model is divided into primary, secondary, and unique features. Features contained in over 66 percent of the DBMS models were classified as essential features. Unique or niche features were characteristics that were contained in less than 33 percent of the DBMS model. The rest were classified as secondary features. As Table 1 indicates, the classified outcomes are as follow:

1) Over sixty-six (66) percent of the DBMS have 13 of the 33 features. These were classified as primary features.
2) Four (4) features are common in between 34 percent and 66 percent of the DBMS. They were classified as secondary features.
3) Sixteen (16) features are found in less than 34 percent of the DBMS so they were classified as niche or unique features.

Figure 3 shows that the features are not equally distributed among the DBMS. More features are contained in less than 34 percent of the DBMS. In other words, many features were supported by only one or two DBMS. OODBMS was also found to provide more unique or niche features than any other DBMS. The outcome of this classification scheme produced an I-Beam schematic model containing thirteen primary, four secondary, and sixteen unique features.

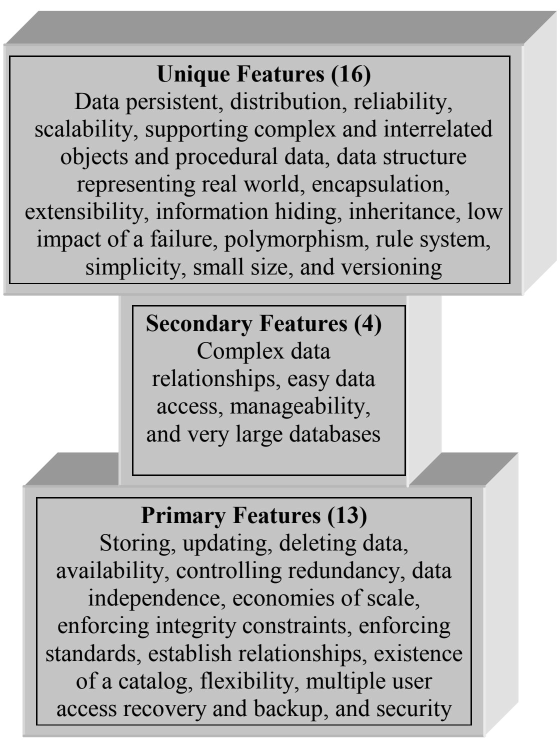

Figure 3: Schematic Model Derived From Literature

Data for each of the database applications were extracted from journal critics, magazine reviews, sales catalogs, white paper, product brochures, the Web sites disclosures, and other online supplements. Thirty of the more popular database applications were identified from the sources indicated. Two products were eliminated from the product list because the vendors did not provide feature descriptions. These database applications are IDS (Integrated Data Store) and Unisys DMS1100 CODASYL. The rest of the data set was then evaluated, tallied, and ranked. Table 2 shows the mapping of the DBMS applications to the primary DBMS model identified.

As can be seen in the second and third column of Table 2, two DBMS have only one application. The two applications and their primary DBMS are Excel, which fits the file processing structure, and IMS, a hierarchy-based application. The rest of the DBMS models have several applications each. Details about each of the models and their applications are presented in Table 2.

Table 2

DBMS Vendors and Products

Vendors	**DBMS Products**	**Type**
Sybase Inc.	Adaptive Server Enterprise 12.5	OR
InterSystems Corp.	Cache	OO
IBM Corp.	IMS DB2 and Database 7.1	H OR
Empress Software Inc.	Empress RDBMS	R
FileMaker Inc.	FileMaker Pro4.5	R
Gemstone Systems, Inc.	Gemstone	OO
Informix Software Inc.	Informix Dynamic Server Universal Server	R OR
IBEX Object Systems, Inc.	Itasca	OO
Computer Associates International, Inc.	OenIngres Jasmine 1.1	R OO
ADB Inc.	Matisse	OO
Microsoft Corp.	Excel Access FoxPro	F.P. R OO
NeoLogic Systems	NeoAccess	OO
Objectivity, Inc.	Objectivity 5.0	OO
Object Design, Inc.	ObjectStore	OO
Oracle Corp.	Oracle 7x Oracle 8i, 9i	R OR
Corel Corporation	Paradox 8	R
POET Software Corp.	POET 5.0	OO
Microsoft Corp.	SQL Server	OR
Cincom systems, Inc.	UniSQL	OR
DBase Inc.	DBase	OO
Logic Arts Ltd	VOSS 3.0	OO
Versant Object Technology Corp.	Versant	OO

Table 3 shows the distribution of features among the listed database applications. The table includes 33 features, the number of the products, and the ranking of each feature presented in descending order.

Table 3

Distribution of the Identified Features of Products

Features	**No. of Products**	**Percentage**	**Rank**
Storing, updating, deleting data	30	100.00	1
Data independence	28	93.33	2
Enforcing integrity constraints	28	93.33	3
Enforcing standards	28	93.33	4
Establish relationships	28	93.33	5
Multiple user access	27	90.00	6
Controlling redundancy	27	90.00	7
Economies of scale	26	86.67	8
Complex Data relationships	26	86.67	9
Easy data access	25	83.33	10
Recovery and Backup	23	76.67	11
Flexibility	18	60.00	12
Security	17	56.67	13
Existence of a catalog	16	53.33	14
Availability	13	43.33	15
Manageability	12	40.00	16
Supporting complex, interrelated objects, and procedural data	12	40.00	17
Data persistent	11	36.67	18
Inheritance	9	30.00	19
Reliability	8	26.67	20
Versioning	8	26.67	21
Distribution	8	26.67	22
Encapsulation	7	23.33	23
Information hiding	7	23.33	24
Polymorphism	7	23.33	25
Rule system	6	20.00	26
Scalability	6	20.00	27
Data structure representing real world	6	20.00	28
Extensibility	6	20.00	29
Very large databases	5	16.67	30
Low impact of a failure size	3	10.00	31
Simplicity	3	10.00	32
Small size	3	10.00	33

Some of the major findings are presented below:

1) Only one feature, storing, updating, and deleting data, is found in 100 percent of the applications.

2) Using the criteria established, 11 of the 33 features were found to be in over 66 percent of the applications, 7 features were found to be common in between 34 percent and 66 percent of the applications, and 15 features were found in less than 34 percent of the products.

Based on the outcomes and the three-tier classification structure, there were 11 primary features, 7 secondary features, 15 niche or unique features. The distribution of the features resulted in an I-Beam structure. The structure and its associated features are presented in Figure 4.

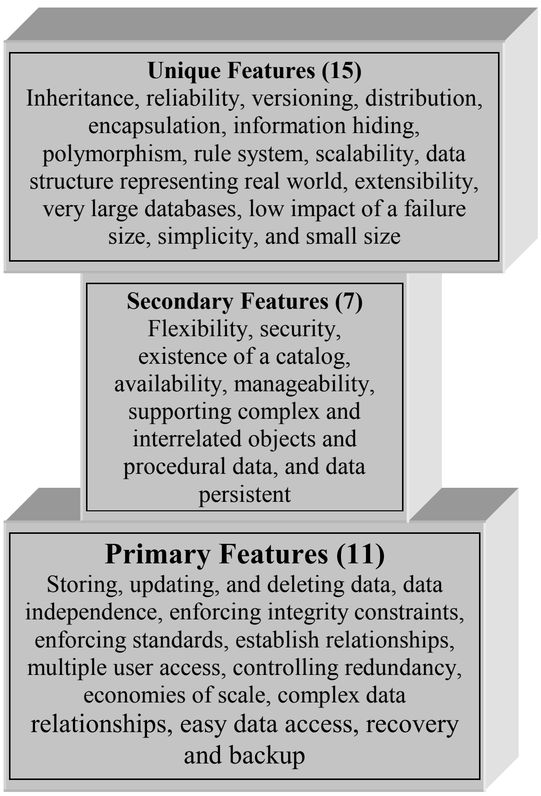

Figure 4: Three-Tier Structure Derived From Product

Table 4 is an assessment of the number of features contained in each of the applications. The features are mapped to the respective tiers to demonstrate the number of primary, secondary, and niche features contained in each of the applications.

Table 4

Products and Number of Features in Each Tier

	Tier 1	**Tier 2**	**Tier 3**	**Total**	**%**
Itasca	11	7	4	22	66.67
Oracle 9i	11	6	13	30	90.91
SQL Server	11	6	12	29	87.88
Adaptive Server Enterprise 12.5	11	6	8	25	75.76
Informix Dynamic Server	11	6	4	21	63.64
DB2 Universal Database 7.1	11	6	1	18	54.55
Versant	11	6	1	18	54.55
Jasmine 1.1	11	6	0	17	51.52
Objectivity / DB 5.0	11	5	1	17	51.52
Oracle 7x	11	5	0	16	48.48
OpenIngres	11	5	0	16	48.48
Universal Server	11	4	9	24	72.73
ObjectStore	11	3	0	14	42.42
Access	11	3	0	14	42.42
Empress RDBMS	11	2	3	16	48.48
Matisse	11	2	0	13	39.39
Gemstone	11	1	13	25	75.76
FileMaker Pro 4.5	11	1	3	15	45.45
POET 5.0	11	0	0	11	33.33
UniSQL	11	0	0	11	33.33
Cache	10	1	9	20	60.61
NeoAccess	10	0	0	10	30.30
Paradox 8	10	0	0	10	30.30
FoxPro	9	4	6	19	57.58
IMS/DLI	9	3	0	12	36.36
VOSS	8	3	4	15	45.45
DBase	6	1	0	7	21.21
Excel	1	0	3	4	12.12
Standard Deviation	3.09	2.39	3.81	6.63	20.10
Mean	9.50	3.23	3.03	15.77	47.78

Some of the major findings about the 28 applications are presented below:

1) Oracle9i has the largest number of total features. Collectively, it is the only application that has over 90 percent of the features.

2) Seven of the applications have over 67 percent of the features. They are Itasca, SQL Server, Adaptive Server Enterprise 12.5, Informix Dynamic Server, Gemstone, Cache, and Oracle 9i.

3) A good number of the applications have identical number of primary features. However, when the two other tiers were also evaluated, they differ greatly.

4) As expected, Excel has the lowest number of features, only 1, even though the manufacturers claimed that they have database capabilities. In reality, this predominantly spreadsheet application actually requires addition DBMS mechanism before it can be considered as a complete database application

4. CONCLUSIONS

This study examined the features that are in six different DBMS models. A schematic model consisting of three levels was used to map the DBMS models. Thirty-three features were identified from the six models. Only one feature was found in all the six DBMS models. That feature was the storing, updating, and deleting feature. Thirteen features were classified as primary, 4 as secondary, and 16 as unique. This distribution resulted in the identification of an I-Beam schematic model.

Using the 33 features and another three-tier classification scheme for assessing the 28 database applications, the results also show that the features formed an I-Beam structure. Eleven of them fall in the primary category, 7 in the secondary category, and 15 in the niche or unique category. One application clearly has more features than the rest of the 27 applications evaluated. This application, Oracle9i, has over 90 percent of the features.

In addition, the results of this study confirmed that the features contained in DBMS models are implemented in some form or another in real world applications. However, the distribution of the features is not equally distributed. They are in the form of an I-Beam structure. The I-Beam structure extracted from the analysis of the DBMS models suggests that the current state of DBMS development is in a continuous state of refinement.

Analysis of the features contained in the applications also showed a similar trend. Even though the distribution of the features extracted from the applications also formed an I-Bean structure, the number of features that fall in each of the categories is different. There are less primary features from the second analysis. These findings suggest that, like in the case of DBMS development, database applications development is also being refined continuously. However, database application developers are focusing on a smaller number of primary features that are critical. In addition, some of the features may mature into the next category as user applications and needs evolve.

Finally, this research study is a relatively comprehensive investigation of current DBMS models and how those features identified in the literature are actually implemented in some of the more popular database applications in existence. As computing technology advances and new form of complex data merge, those six DBMS models may have to be revisited. Some contemporary and innovative features may need to be added, others removed because they are no longer necessary, and still others refined to better meet the needs of users and the changing environment. One such opportunity that may need scholars and developers to begin dialogue about the next generation of DBMS is already in existence. The Internet and the proliferation of ubiquitous computing will greatly impact the development of the next generation of DBMS. In addition, the increasing demand for faster, larger, more dynamic, and intelligent multinational data applications will also bring about the need for re-examining the inadequacies of the DBMS models in existence. This study can be used as the basis for such future work. The schematic models as well as the extracted structures presented here are examples of flexible approaches that can be replicated or adapted for future research into the development of database theory. Some of the material presented in this study, particularly those on DBMS models and the features, can be used as instructional material for the teaching of DBMS history and conceptual evolution.

5. ACKNOWLEDGEMENTS

We would like to thank Dr. Sathi Mahesh at the University of New Orleans for his constructive suggestions on this research.

6. REFERENCES

Apicella, M., 1998, Object Databases. LanTimes [On-line]. Available: http://www.lantimes.com/testing/98may/805a057a.html.

Burleson, D., 1999, Inside the Database Object Model. CRC Press, New York.

Tao, Y. and W. Grosky, 2000, "Image Indexing and Retrieval Using Object-based Point Feature Maps."

Journal of Visual Languages & Computing, 11 (3), pp. 323-343.

Chung, M., and et al., 2001, "Querying Multiple Data Sources via an Object-Oriented Spatial Query Interface and Framework." Journal of Visual Languages & Computing, 12 (1), pp. 37-60.

Connolly, T. and C. Begg, 1999, Database Systems. Addison-Wesley, New York.

El-Rewini, H. and S. Hamilton, 1995, "Object Technology." Computer, 28, pp. 58-59.

Gartner Group, 2000, "Overview of 1999 DBMS Market Size, Vendor Shares and Forecast." Gartner Group [On-line]. Available: http://www.gartner.com/webletter/progress_software/article2/article2.html

Grimes, S., 1998, "Modeling Object/Relational Databases." DBMS [On-line]. Available: http://www.dbmsmag.com/9804d13.html

Larson, J., 1995, Database Direction. Prentice Hall PTR, New Jersey.

Muller, R., 1999, Database Design for Smarties. Morgan Kaufmann Publishers, California.

Teorey, T., 1999, Database Modeling & Design. Morgan Kaufmann Publishers, California.

Vandenbussche, J. and J. Paredaens, 1995, "The Expressive Power of Complex Values in Object-Based Data Models." Information and Computation, 120 (2), pp. 220-236.

Yazici, A. and R. George, 1999, Fuzzy Database Modeling. Physica-Verlag Heidelberg, New York.