

The Database Plantation

Victoria E Holt

The concept of a database is not only about the storage of data but it is about the development of the database, the global economy, trends for the latest technology, delivering and sharing reliable data in a managed fashion. This eclectic concept tends towards a systemic approach to the complex database. A new way of thinking about the database system is that of a plantation.

A plantation is an intentional planting of trees and shrubs usually associated with economic models. Multiple numbers of employees and types of machinery carry out the work, as in the setting up, use and monitoring of a database. There are various different types of plantation designed to produce different outputs. Industrial plantations aim to produce high volumes of wood in a short time such as Christmas trees, paper, teak; environmental plantations are set up to protect the soil, watershed or to reduce carbon dioxide; whilst some produce high value food crops such as tea, coffee, cacao, orchards, fruit, tobacco, sugarcane, cotton, rubber and rice. This can be compared to the multiple purposes of databases. The numbers of plantations, as with databases, are set up to serve the demand for certain products/services.

As with anything the plantations need to be managed to deal with cultivation, growth cycles, harvesting, rotation period and the process is iterative. There are various issues which affect the plantation such as tree species that do not naturally occur in the area, the plantations are young forests, monocultures designed for high yields. In some countries social problems have occurred where natural forests have been converted to plantations. This causes problems with biodiversity, loss of habitat if non-native plants are grown or it can be a benefit if the land was degraded before or they can create corridors for wildlife and act as a buffer for native forests. The environment has a large impact on the plantations with global warming, the weather, tree stock, GM crops, insect infestations, globalization etc. Thus it can be seen that the problems of achieving successful plantations, can be compared to the nature of monitoring successful data bases.

The plantation is a metaphor for the database. The description and effects match perfectly and help holistically identify the system. A database is a living system which is always evolving, reflective with multi and mutual causation.

The plantation metaphor is summarised in figure 1

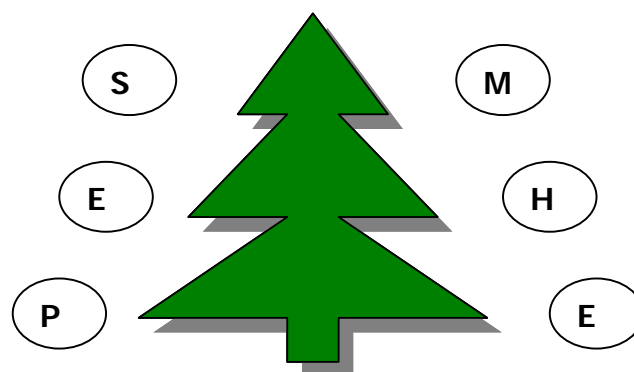


Figure 1 The plantation metaphor

Key

S – Species	M – Managed
E – Environment	H – Harvesting
P – Planting	E – Economy

In comparison the database is created by design with an aim of providing and storing a certain set of data. There can be team or sole developers. The database can be used to hold the financial data of the banks and insurance companies, shop goods, research data such as climate change statistics, warehouse stock, medical records, government policies, a document store use for internal secured applications or for web access. The number of databases will increase with the demand for storage, for new IT projects and new technology. A database grows by design as more data is input or organically as extra bits of functionality are added over time. The growth sometimes goes in fits and bursts or regularly the data is reduced in size as regular maintenance archives data or people download the content. An iterative process. There are many database applications, these are but some

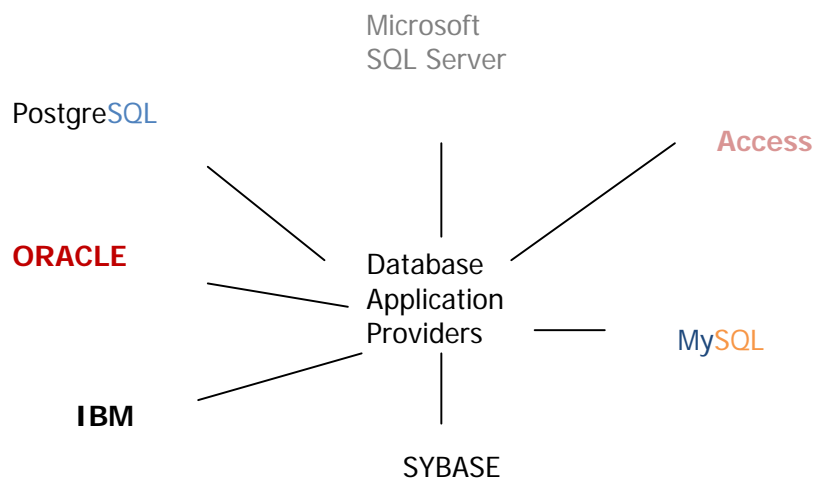


Figure 2 Some of the major Database Application providers

To name but a few of the Database Management Systems (DBMS) (design feature of M) out there see figure 3

Database Software	Organization	First public release date	Software
Microsoft SQL Server	Microsoft	1989	Proprietary
Oracle	Oracle Corporation	1977	Proprietary
DB2	IBM	1982	Proprietary
MySQL	MySQL AB	November 1996	Open Source
PostgreSQL	PostgreSQL Global Development Group	June 1989	Open Source
Sybase	Sybase	1988	Proprietary

Figure 3 table of a few DBMS systems (Feature M)

Proprietary software is software owned by a single organization or individual. In contrast with open source software which refers to software that is created by a development community rather than a single vendor.

Microsoft SQL Server is for enterprise level users offering high availability solutions for medium size databases. Formerly named Yukon the revolutionary version of Microsoft SQL Server 2005 provides vast range of new features and tools. The enhancements cover new tools for the database administrator and features within the database server which allow scalable and high performance database applications to be built for mission critical software. On the development side the integration of the .NET framework into SQL Server allows more complex retrieval and manipulation of data with greater efficiency. In the business intelligence (BI) area it is possible to analyze and manipulate data to allow future trending and pattern analysis to take place and then to be provided with the reports from within the database server. There are various cut down versions of this database management system to aid in the development of these applications to providing mobile and systems with less functionality.

Oracle is an enterprise level database management system which is available on multi-platforms. Oracle tends to be used by the higher end of the market range due to the cost of Oracle being greater than that of its competitors. Oracle is widely used due to its reliability and ability to store vast amounts of data. The data retrieval can often be slower than other systems but the ability to retrieve this volume of data is advantageous. The current version of Oracle 10g is professed to be the world's first self-managing database which is aiming to deliver faster performance and higher availability with smaller management costs. The strategy has two goals "simplify the management of the database itself, by automating much of the monitoring and maintenance, and provide data centres with a broadly focused , rich management tool that will allow them to manage all the components - not just the database – that the data centre must manage, regardless of where the components are (in the data centre,

deployed across a grid, scattered worldwide) and what they are (storage, clusters, application servers, and so on). " Kelli Wiseth, Oracle

Oracle's aim is the holistic self managing database which currently has its ability to diagnose its own performance problems. This design will allow the existing hardware to be fully utilised. In the long term it is looking at complete automation so that the database shouldn't require any administrative intervention with the loss of complexity but still retain the flexibility. The entire system lifecycle is supported by the grid management feature.

Oracle 10g has a new tool grid computing which aims to reduce costs and ensure acceptable performance, and infinite scalability. Grid computing takes its name from the electric power grid. It is a way of sharing computer resources on demand within a tightly coupled network. It has two goals, higher availability and the ability to add resources on demand. The Search For Extraterrestrial Intelligence (SETI) project is a well known example of distributed grid computing. This arose from the need to solve tasks with highly-parallel computational problems beyond a computer to solve. The master computer issues small subprograms to independent servers which work independently to return results back to the master computer. Oracle works in a way that involves direct hardware allocation and the allocation of remote servers to execute Oracle tasks. Enterprises can then share computing, storage, data, programs, and other resources in a dynamic fashion.

MySQL is an open source database which has often been considered the hobbyist-type database. MySQL has been seen as a toy database in the past even though it was capable of working in many environments and supporting large loads. MySQL has three software priorities good performance, reliability (have as few as bugs as possible) and ease of use using a fifteen minute rule. A rule that means it is possible to download, install and have MySQL database up and running in just fifteen minutes. MySQL 5 is the most important release ever as the new version now has many more features which it had previously lacked such as stored procedures, views, data dictionary, triggers, information schema which makes the database become an enterprise grade database.

A key feature of MySQL is the various pluggable data storage mechanisms such as MyISAM, InnoDB etc and Falcon the next generation transactional engine. There are many other storage mechanisms available within the product so it is necessary to plan which storage engine to use depending on what kind of systems is under development.

The PostgreSQL relational database system (RDBMS) came from the POSTGRES project at the University of California at Berkeley where it was developed from the object relational database management system research project. The object relational database or object relational database management system is a relational database management system that allows developers to integrate their own custom data types and methods into the database. It is also an open source database.

IBM DB2 is another enterprise level database. IBM DB2 Universal Database supports all known platforms. The DB2 Universal Database (UDB) is a cross-platform database system developed by IBM. Development started on mainframe platforms and in 1996 announced a version for distributed platforms. With this version, DB2 was able to store all kinds of electronic data, including traditional relational data, as well as audio, video, and text documents and it was the first version optimized for the Web.

Sybase ASE started life as Sybase SQL Server. Microsoft was a Sybase distributor who bought the source code of Sybase SQL around 1994 renaming that product Microsoft SQL Server. Sybase and Microsoft have been developing their products independently and Sybase has been focused on maximizing performance and reliability, and catering to the high end of the OLTP market.

So when deciding the criteria used to selecting the database platform required it becomes necessary to examine what features and availability of service is required by the specification. Each database has it's own merits and features which stand out. The environment affects the database server with the demand for users, corporate funding, viruses, the government, legislation world wide web developments etc. Each database generally has a limited life expectancy due to technology progressing.

The life of the plantation and the life of the database are intertwined. The metaphor helps to visually represent the complex system we are dealing with. The stages of the database plantation are:-

P – Planting

The database is created by design and a need to provide a certain set of data. This engaging with the problem situation requires modelling the database with ethical database design and development. This full database design needs to be fully documented. The tree growth land management can be compared to release management and actual storage of the data required.

E – Environment

The environment affects how databases evolve over time. If you think of ecological impact this is equivalent to a sudden change in technological advances security breaches. The environment affects design and what information is required to know the situation. A data quality baseline, data definition, metadata definition and governance methods are affected and influenced by the current environment. Regular health checks are required to ensure the database stays within the set guideline like keeping the weeds under control or the possible deforestation of plantations.

S – Species

There are many database providers which provide different functions dependant on the type of data to be stored. Some are just for basic data, some for security details and some for web content and some for internal systems. Record keeping and long term archiving with tool identification is compared to keeping seeds for future pollution to stop species becoming extinct and so that the trees stock is of good quality and not affected by things such as air pollution. MySql has a pluggable data storage mechanism which equates to a crop rotation cycle.

M – Managed

All of the different management systems, some of which are outlined above, need to include the following. A database like a plantation needs to be managed through every step of the growth cycle, ensuring backups are performed, ITIL methods are followed, housekeeping done, change management carried out to make sure the database is always available and managed. Regular auditing is required to make sure everything is under control. A key aim of Oracle is the self managing database which shares physical resources.

H – Harvesting

The data is delivered through whatever format to the customer at regular intervals or on request and may be compared to the harvest. The fruits of the harvest could be available any time during 24/7. The mission critical performance baseline, high availability databases can be compared to high volume crops with convenient transport accessible.

E – Economy

The finance cost of how much it costs to build the database, work on the plantation, support the data and manage it is always of prime concern. The cost can be recorded and analysed using business intelligence software. This ability to do trend and pattern analysis helps the business become more profitable and meet the demands of the consumers. MySQL have a 15 minute rule which means there is not a long lead time between deciding on the product and implementation which reduces the cost as well as the software being open source.

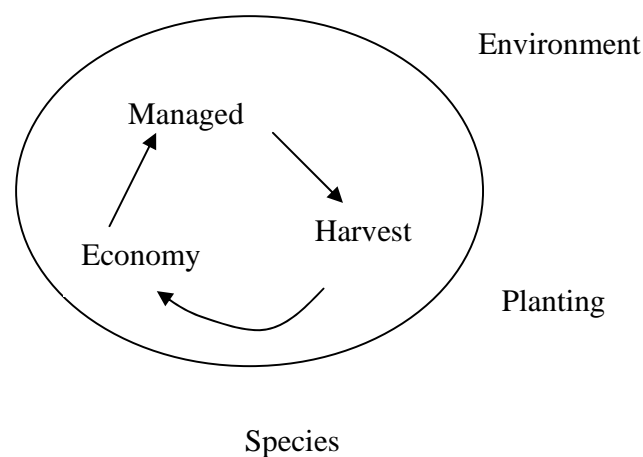


Figure 2 Database survival system

The database plantation metaphor helps describe this system in terms of a database survival system as seen in figure 2. This shows the effect each layer of the plantation

has on another and thus each layer of the database on another. Each part of the database system is interconnected.

The Plantation is highly complex but yet adaptable depending on the environment or finance. There are some key feature amongst the databases which are driving the way forward and will be key features for future development within the database arena. These are:-

- Self managing database
- Pluggable storage engines
- Trend and pattern analysis
- Ease of use
- Sharing data
- Advanced security
- Data quality

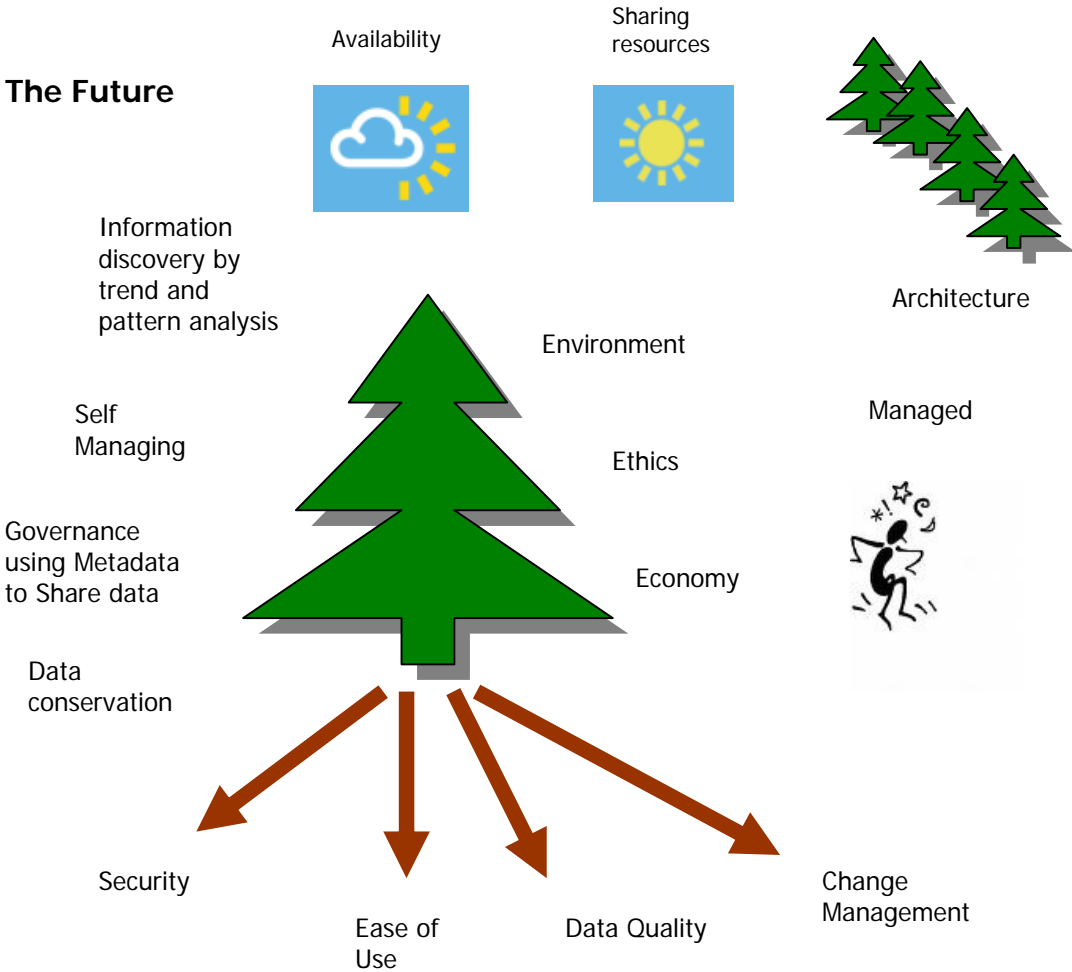


Figure 3 The Database Tree of knowledge showing overall design considerations.

Resources

<http://www.postgresql.org/>
<http://www.oracle.com/index.html>
<http://www.mysql.com/>
<http://www.microsoft.com/sql/default.mspx>
<http://www.sybase.com/>
<http://www-306.ibm.com/software/data/db2/>

Comparison of relational database management systems
http://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems

Features, strengths and weaknesses comparison between MS SQL 2005 (Yukon) and Oracle 10g databases
http://www.wisdomforce.com/dweb/resources/docs/MSSQL2005_ORACLE10g_compare.pdf

Oracle Database 10g, The world's first self-managing, grid-ready database arrives by Keith Wiseth <http://www.oracle.com/technology/oramag/oracle/o3-sep/0531ogcover.html>

PostgreSQL vs. MySQL vs. Commercial Databases: It's All About What You Need
<http://www.devx.com/dbzone/Article/20743>

A quick introduction: What is Sybase ASE ?
http://www.sypron.nl/whatis_ase.html

Introduction to DB2 UDB
<http://www.phptr.com/articles/article.asp?p=375536&rl=1>

Grid computing
http://www.webopedia.com/TERM/g/grid_computing.html

Expert advice: Grid computing and Oracle10g By Brian Peasland, SearchOracle.com Expert
12 May 2004 | SearchOracle.com
http://searchoracle.techtarget.com/originalContent/0,289142,sid41_gci964014,00.html

Oracle Dynamic Grid Computing
by Donald K. Burleson
http://www.dba-oracle.com/art_dbazine_grid.htm