Where from? Where now? Where to?

Dr Alvaro A A Fernandes, Department of Computer Science, University of Manchester, discusses database technologies…

Database management systems (DBMSs) have been remarkably successful in fulfilling a host of challenging expectations that both researchers and practitioners have projected upon them. To most of such challenges researchers have found elegant, robust and durable solutions. Nonetheless, the dynamic nature of society at large, and of the information technology landscape more specifically, creates pressures to which database technologies must respond by adapting and evolving.

DBMSs are sophisticated and highly evolved software systems. Their mission is to support users and applications in storing and retrieving information conveniently and efficiently. To understand where database technologies have come from, where they are at present, and where they are heading in the short and in the long term, we need to look more closely at that mission statement.

Where from?

Database technology arose, a few decades ago, out of the need to respond to the opportunities created by mass storage systems. It soon became clear that storing ever larger volumes of data was, and would remain, cost-efficient. Database research responded with specific, sophisticated principles, methods, techniques and tools that allowed the data assets of organisations to be conveniently and efficiently stored and retrieved.

Convenience required, for example, that application programs enjoyed transparency with respect to the physical aspects of the storage media. The development of high level, declarative query languages such as SQL was another response to the need for convenience – in this case, of retrieval rather than of storage – insofar as the alternative would have been for programmers to continue writing one-off, costly and error-prone retrieval procedures with consequent risks to reliability and scalability.

Efficiency required, for example, that the storage of data is complemented by storage of both descriptive information (for example, metadata such as type information, value distributions, etc.) and auxiliary structures (eg. indices that underpin the use of efficient methods for accessing and updating the data). The development of sophisticated query optimisation techniques to select the best access paths automatically and reliably is another example of an extremely successful response.

In the last decade or so, societal and technological changes have so re-shaped computing environments that, like many other areas in computing, database technologies have had to evolve quickly. The pressures that have caused DBMSs to change and adapt are quite diverse and complex, but we categorise them here into two broad, composite kinds: those to do with diversity and quantity, and those to do with integration and distribution.

Where now?

Pressures to do with diversity and quantity include those resulting from increasing type heterogeneity, more diverse data models, and the need to transform raw data into actionable information.

Until very recently, DBMSs were only expected to store simple, primitive types such as numbers and simple strings. However, the public sector, for example, would benefit from the ability to store and retrieve conveniently and effectively digitised map information or documents. Local health authorities might be interested in tracking regions with a high incidence of allergic disease lying adjacent to contaminated sites. Planning departments might benefit from being able to manage linked collections of documents that represented stages through which some policy initiative had been shaped by public consultation events.

Organisations would also benefit from the ability to extract from the vast data stocks they have amassed information that was good enough to be acted upon, or even to inform or shape policies and strategies. For example, a local health authority analyst might wish to explore, interactively, the impact of decontamination initiatives on the health needs of the population over time and location. Likewise, decision-makers would greatly benefit from the ability to generate, from the available data, a decision model as to where, when and how to decontaminate a given site on the basis of predicted patterns of occupancy in surrounding areas.

In response to type heterogeneity, database researchers have worked to extend the range of types that a

EPSON

Use EPSON projection technology to promote interactivity and develop a co-operative approach to learning. www.epson.co.uk

DBMS can support. For example, most commercial DBMSs have primitive support for storing images or sounds. Much more challenging is the ability to respond to the need for spatial and historical information to be stored conveniently and efficiently. To achieve this, core aspects of DBMSs must be changed in non-trivial ways. Mainstream DBMSs have yet to rise to this challenge, but they do often offer add-on components that ameliorate matters, whilst research presses forward with the task of overcoming the core foundational issues.

The widespread adoption of XML as the lingua franca for document-based information interchange is indicative of the need to respond to more diverse data models. Again, mainstream DBMSs have been trivially extended with support for XML, but the loss of convenience and efficiency that results from simplistic responses has caused researchers to embark on the task of overcoming the core foundational issues involved. Two different strands of activity are ongoing: one on query languages for XML (of which XQuery is the most likely candidate standard) and another on XML-native DBMSs, in which the core DBMS functions are not implemented in terms of the structured, (object) relational data model underlying SQL but rather, from scratch, as a semi-structured data model in which graphs, rather than tables, are the most fundamental type abstraction.

As to the need for conveniently and efficiently transforming data into actionable information, most mainstream DBMSs have already capitalised on research in data warehousing and data mining with different degrees of convenience and efficiency. The full benefits of both technologies are still to be reaped. The same strategy of adding extra components and layers has resulted in DBMS products that offer limited online analytical processing and limited scope to generate descriptive and predictive models from stored data. More research is needed on foundational issues, but here again, there are grounds for optimism.

Our research group in Manchester has developed a prototype for a new class of database systems, called combined inference database systems. Their main benefit is that they seamlessly integrate querying the data and mining it, in a manner that enhances, rather than curtails – as 'bolt-ons' are prone to do, the convenience and effectiveness of the DBMS.

Pressures to do with integration and distribution are a side effect of the paradigm shift to net-centric computing and the consequent explosion in the number and variety of potential data resources that have become available to organisations.

Until very recently, DBMSs were centralised systems built under the assumption that they would run over local resources over which organisations had full administrative control. This assumption is no longer valid, now that the network has become ubiquitous and transparent: both data and computational resources can be anywhere and are not necessarily under (or amenable to) central administrative control. Current DBMSs have made modest progress towards delivering convenience and efficiency in response to challenges related to integration and distribution. For example, there are commercial DBMSs that allow an organisation to define a globally valid description of autonomous, local data resources and evaluate queries over the latter. However, the mechanisms used are not standardised.

Ideally, organisations would prefer to reduce the inconvenience and inefficiencies resulting from the ad hoc nature of currently available facilities by using standard reference implementations of such advanced functionality. In collaboration with colleagues from Newcastle and Edinburgh (as well as IBM UK), our research group in Manchester has been actively steering the emerging standards for service-based data access and integration in the web and in the Grid. In particular, we have developed high level service-based middleware for the UK e-Science programme that allows, on the one hand, heterogeneous, autonomous, distributed data resources to be accessed homogeneously and, on the other, for queries to be executed over implicitly parallel, distributed computational resources.

Another source of pressure relating to distribution is streaming data. The public sector, for example, might have an interest in systems that better identify traffic patterns over space and time using data streaming from GPS-enabled devices such as palmtops or mobile phones. Likewise, data might stream from sensor networks over river basins and enable better flood protection measures. Here, very modest progress has been made, but more substantial advances are in the pipeline.

Where to?

Database technologies will no doubt improve in response to the challenges cited and new ones that are sure to arise. As research results make their way into commercial DBMSs, these systems will continue the tradition to deliver convenience and efficiency in the production of actionable information from heterogeneous, autonomous, distributed resources abiding by more diverse data models with increasing levels of network and resource transparency.

'Organisations would also benefit from the ability to extract from the vast data stocks they have amassed information that was good enough to be acted upon…'

THE UNIVERSITY of MANCHESTER

Dr Alvaro A A Fernandes
Lecturer
Information Management Group
Department of Computer Science
University of Manchester
Manchester M13 9PL
Tel: 0161 275 6199
Fax: 0161 275 6236
a.fernandes@manchester.ac.uk
img.cs.man.ac.uk

EPSON

Use EPSON projection technology to promote interactivity and develop a co-operative approach to learning. www.epson.co.uk