Its my pleasure to announce to the formation of the Data Engineering Workgroup on Self-Managing Database Systems (DEW-SMDB), a formal workgroup of the IEEE Technical Committee on Data Engineering (TCDE). http://db.uwaterloo.ca/tcde-smdb/

To advance research and development in self-managing database and information management systems the workgroup will sponsor workshops in conjunction with the ICDE, foster publications, and maintain a collection of key resource links on its website. I am pleased that our initial executive includes some of the leading advocates of self-managing database technology from across geographies in both industry and academia.

Alfred Whitehead, the preeminent American mathematician once wrote that "Civilization advances by extending the number of important operations which we can perform without thinking about them." Database technology and in particular enterprise scale relational database technology, has reached a milestone in advancement of feature rich capabilities and massive data volumes which compels a quantum improvement in the administration profile of these systems.

Since the development of the relational model by E.F. Codd at IBM, relational databases have become the de facto standard for managing and querying structured data. The rise of the Internet, online transaction processing, online banking, and the ability to connect heterogeneous systems have all contributed to the massive growth in data volumes over the past 15 years. Terabyte sized databases have become commonplace. Concurrent with this data growth, have come dramatic increases in CPU performance, spurred by Moores Law, and improvements in disk fabrication which have improved data density limits for persistent disk storage. Relational databases lie behind 70% of the world's structured data, and much of the world's unstructured and semi-structured data is evolving into structured management systems as well. Practically everywhere data is processed in significant volume there is an information management system. To help manage the massive volume of data, and exploit the available hardware, modern Relational Databases Systems (RDBMSs) have been enhanced with powerful features to maximize efficiency and availability, providing numerous topology options for structuring the data and the storage, for indexing and caching, and for managing system workloads. These features are much needed in environments where systems can contain hundreds of disks and CPUs, massive volumes of data, and are often required to run 24 hours a day, 7 days a week (24x7).

The powerful capabilities of modern RDBMSs add complexity to the system too however, and have given rise to an entire professional domain of experts, known as Database Administrators (DBAs), to manage these systems. The problem is not unique to RDBMSs, but is ubiquitous in technology today. Some have called this explosive growth of complexity creeping featurism. In fact many companies now spend more money recruiting administrators to manage their technology than the money they have spent on the technology itself.

The only real viable long term strategic solution is to develop systems that manage themselves. Such systems are called self-managing or autonomic systems.

The following challenges in information management require the industrys attention:

Self-Configuring refers to function that connects and configures components so that technology is operational with minimal human intervention. For database servers this can include server and storage topology configuration, installation, initial database design, and configuration of the database server (there are several dozen possible options). Self configuring also includes the ability for a database server to be easily deployed, configured and interconnected with other middleware and applications.

Self-Healing refers to the ability of a system to detect problems or problem trends (which have not yet manifested as problems) and either take corrective action, or provide advice and automated notification to an end user who can then take corrective action. With all the sophistication and redundancy in database systems, things can and will occasionally fail, which makes the automated problem determination aspect of self-healing a critical domain. In automated problem determination, a self-managing system detects problem trends (such as disk storage is filling up at a detectable rate, from which storage constraint can be projected), or detects and notifies administrators when sudden errors arise.

Self-Optimizing is the domain of system performance tuning. For databases this includes tuning of subtle I/O and caching parameters, memory distribution, and physical database design. Physical database design is the broad area of database administration that deals with the selection of attributes that structure and control the way data is physically stored. Some common physical database design problems include: determining ways to cluster (sort) data on disk, which indexes (fast lookup structures, such as B+ trees) should be created, and what subset of the data should be pre-computed or replicated (materialized views).

Trust in self-managing systems. Even as we advance the underlying technology of self-managing systems, we need to concomitantly advance the trust of the administrators and executives who depend on these new self-managing capabilities. Building trust is as much to do with emotional invest and risk assessment as much as with technology.

Benchmarking of self-managing information systems. How can we objectively assess the success of data management systems to be self-managing? Just as benchmarks have been created in other domains to help assess the maturity and performance of system capabilities (such as database query or transaction performance) the development of industry standardized benchmarks to evaluate attributes of self-management will further compel the industry to compete for TCO, just as it has done historically for system performance.

Self-Protecting systems have the ability to anticipate, identify and protect against security attacks. This scope varies from static defenses such as virus protection and data encryption to more advanced dynamic defenses such as behavior analysis where user/application activity can be analyzed for negative behaviors, tracked, and adaptively undone.

System-wide self-management. Finally but not least, the holy grail of systems administration comes through not only creating piece-wise self-managing components (such as a self-managing database management system) but also by enabling large complex stacks of hardware and software in a truly cooperative self-managing system. This cooperative notion can only be achieved through the development of standards and multi-component technology.

The future evolution of these technologies will lead to self-managing systems that are require a minimal amount of human interaction, and a maximum amount of efficiency, meaning that information management systems will become both easy to administer and resource efficient.

Its our pleasure to announce the formation of this workgroup, and we invite everyone interested to follow our web site for news and announcements.

"Any intelligent fool can make things bigger and more complex... It takes a touch of genius - and a lot of courage to move in the opposite direction." - Albert Einstein