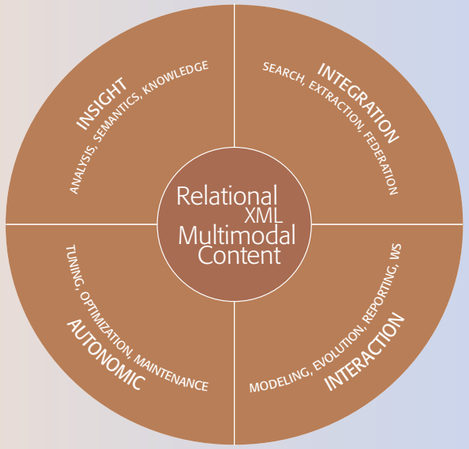
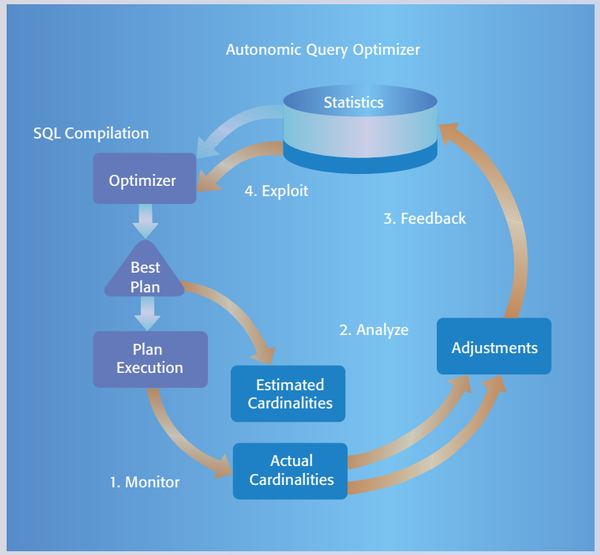
For more than three decades, IBM Research has been a leading innovator in data management. In 1970, E. F. Codd published the landmark paper that established the relational model that underlies most of today's database systems. In 1973, work was started on the seminal System R Relational Database Management System (RDBMS) that provided the first implementation of the Structured Query Language (SQL) as well as development of critical RDBMS technologies for query compilation, cost-based optimization, ad hoc query formulation, and online data definition. In the ensuing years, additional key innovations followed, such as the Algorithms for Recovery and Isolation Exploiting Semantics (ARIES) family of transaction locking technologies, the Starburst framework for extending RDBMSs to handle new forms of information, the Garlic approach for data federation across diverse data management systems, and the Query By Image Content (QBIC®) system.

IBM's current directions in database research address emerging requirements in several new areas, including scaling RDBMSs based on grid computing paradigms, providing DB2® support for standards-based Web services and the IBM e-business on demand business model, and extending IBM's DB2 Universal Database™ system to handle XML content using the powerful XML Query (XQuery) language. In addition, our research program has expanded to explore novel approaches for making database systems self-managing or autonomic, integrating information across heterogeneous data sources, including support for real-time event processing and data streaming environments, as well as extracting knowledge from structured, unstructured, multimedia, and sensor data.





AUTONOMIC DATABASE SYSTEMS

Autonomic computing techniques are being investigated for creating self-managing database systems that simplify and automate many of the tasks associated with database administration. The overall objective is to make RDBMSs self-configuring, self-healing, self-optimizing, and self-protecting, thereby reducing the total cost of ownership and improving quality-of-service. This multi-disciplinary effort combines database research, artificial intelligence, information theory, statistics, and control theory. Researchers are exploring a wide range of topics in self-managing databases, including developing advisor tools to reduce human effort in database design, techniques for optimizing queries during execution, and algorithms for mining diagnostic information to determine the cause of the problems as well as the solution.

INFORMATION INTEGRATION AND KNOWLEDGE EXTRACTION

The continuous expansion in both the volume and diversity of data information makes it increasingly difficult for enterprises to effectively use this data in business decisions. Enterprises need to deal with not only structured data from RDBMSs, but also with unstructured content such as text (e.g., e-mails, Web pages), audio (e.g., call center recordings, meeting logs), and video (e.g., corporate broadcasts, learning content). Data is also increasingly distributed across systems and organizations making data federation an important new problem. By requiring analysis across all of the data modalities, extraction of meaningful intelligence in this environment goes beyond traditional OnLine Analytical Processing (OLAP) and data mining. Research in information integration and knowledge extraction is being pursued to enable enterprises to more effectively harness distributed, heterogeneous data sources. Important research directions include novel uses of XML, metadata, schema mapping technologies, and ontologies to aid in the syntactic and semantic integration of information, as well as the use of Web services and grids for connecting and sharing processing among distributed database systems.