AUTONOMIC COMPUTING

IBM product review (Courtesy of IBM)

RESEARCH FOCUS

Over the past forty years the computer industry has been defined and driven by an obsession with faster, smaller, and more powerful.

It's time to re-examine the goal.

The focus has been on raw processing power and the individual components that allow ever smaller and greater capacity to store, process and move data. And while scientists and researchers have met this demand with astonishing regularity, we have missed an opportunity to look at the evolution of computing from a more holistic perspective.

There are a number of immediate needs that require us to adjust our thinking and reinterpret our approach to computing in general, and specifically to the interaction between computer hardware, software and networks. The current strain on I/T services demands that we turn our best minds to developing computer systems to be more automated and less reliant on human intervention.

IBM calls the next generation of computing autonomic,▪ because it must act like our autonomic nervous systems.▪ It must provide an unprecedented level of self-regulation while hiding complexity from the user. And it will be a radical shift in the way we conceive and develop computing systems today. This will call for more than retooling old systems -- autonomic computing▪ calls for a whole new area of study.

We're committed to making this happen. In fact, we've reorganized our Research division around this goal.

IBM invites you to join this effort. To encourage involvement with the rest of the research community, IBM plans to host and sponsor conferences, develop an Advisory Board and begin funding research on autonomic computing through awards and fellowships. Please read on to learn more about the benefits, challenges and on-going research relating to autonomic computing.

Research into creating autonomic▪ systems won't be easy, but future computer systems will have to incorporate increased levels of automation if we expect them to manage the ballooning amount of data, the ever-expanding network and the increasing might of processing power.

To create autonomic systems researchers must address key challenges vith varying levels of complexity. Here is a partial list of the challenges we face.

- System identity: Before a system can transact with other systems it must know the extent of its own boundaries. How will we design our systems to define and redefine themselves in dynamic environments?
- Interface design: With a multitude of platforms running, system administrators face a briar patch of knobs. How will we build consistent interfaces and points of control while allowing for a heterogeneous environment?
- Translating business policy into I/T policy: The end result needs to be transparent to the user. How will we create human interfaces that remove complexity and allow users to interact naturally with I/T systems?
- Systemic approach: Creating autonomic components is not enough. How can we unite a constellation of autonomic components into a federated system?
- Standards: The age of proprietary solutions is over. How can we design and support open standards that will work?

- Adaptive algorithms: New methods will be needed to equip our systems to deal with changing environments and transactions. How will we create adaptive algorithms to take previous system experience and use that information to improve the rules?

The pursuit of autonomic computing gives us a unique opportunity to define the next era of computing.

Since their conception, computers have failed to meet up to expectations nearly as often as they have exceeded them. Autonomic computing provides an opportunity to recalibrate the higher purpose of computers from one of convenience, to a tool that allows us to exploit previously un-navigable challenges.

Progressively autonomic computers will enable the tools to analyze complex problems. For instance, machines with cellular architecture, such as Blue Gene, will enable the study of phenomena occurring in fractions of a second at an atomic scale. Greater access to more computing power through grid computing combined with the implementation of open standards will enable researchers to more easily collaborate on complex problems for the global good. Autonomic computing will be able to better harness existing processing power to run complex mathematics for functions such as weather simulations and other scenarios that inform public systems and infrastructure.

The Grand Challenge of autonomic computing gives us a new horizon to reach for.

IBM is defining a path to this next era, and now looks to the scientific community to join us in making it a reality. In the following months, IBM will be working jointly with individuals and institutions to provide funding in the form of awards and fellowships to empower the technical community to meet these challenges.

The realization of the autonomic computing vision will require the energy and resources of researchers and labs around the globe. IBM Research is committed to working both within its own global labs and with researchers and engineers worldwide to accomplish this goal.

External Projects

- Berkeley University: OceanStore
- Berkeley University: Recovery-Oriented Computing
- Carnegie Mellon University: Self-securing Storage
- Columbia University: Autonomizing Legacy Systems
- Cornell University: Astrolabe
- Georgia Institute of Technology: Qfabric

IBM Research Projects

- Discovering Dynamic Dependencies in Enterprise Environments for Problem Determination
- Dynamic Surge Protection: An Approach to Handling Unexpected Workload Surges With Resource Actions That Have Dead Times
- Eos: An Approach of Using Behavior Implications for Policy-based Self-management
- Dynamic Surge Protection: An Approach to Handling Unexpected Workload Surges With Resource Actions That Have Dead Times
- Generic Online Optimization of Multiple Configuration Parameters With Application to a Database Server
- LEO -- DB2's Learning Optimizer
- Policy-Based, Autonomic Storage Allocation Manager
- Storage Tank

Projects outside IBM

Work contributing to the creation of autonomic computing systems reaches beyond IBM's laboratories. The links below highlight projects underway at universities contributing to autonomic computing or related fields. As the I/T community begins to redefine its goals to meet the needs of our changing world, this list will grow.

If we should be linking to your work, send us a note with a link to your project information and a brief description of what you're doing. Send this information to webmaster@watson.ibm.com .

Berkeley University of California: OceanStore
OceanStore is a global persistent data store designed to scale to billions of users. It provides a consistent, highly-available, and durable storage utility atop an infrastructure comprised of untrusted servers. Any computer can join the infrastructure -- users need only subscribe to a single OceanStore service provider, although they may consume storage and bandwidth from many different providers.
Contact: John Kubiatowicz is a researcher at Berkeley exploring the space of Introspective Computing, namely systems which perform continuous, on-line adaptation. Applications include on-chip tolerance of flaky components and continuous optimization to adapt to server failures and denial of service attacks.

Berkeley OceanStore homepage

Berkeley University of California: Recovery-Oriented Computing
The Recovery-Oriented Computing (ROC) project is a joint Berkeley/Stanford research project that is investigating novel techniques for building highly-dependable Internet services. ROC emphasized recovery from failures rather than failure-avoidance. This philosophy is motivated by the observation that even the most robust systems still occasionally encounter failures due to human operator error, transient or permanent hardware failure, and software anomalies resulting from software aging.
Contact: David Patterson is a Professor in Computer Science at UC Berkeley working on the ROC project.

Berkeley ROC project page

Carnegie Mellon University: Self-securing Storage & Devices
Self-securing storage is an exciting new technology for enhancing intrusion survival by enabling the storage device to safeguard data even when the client OS is compromised. It capitalizes on the fact that storage servers (whether file servers, disk array controllers, or even IDE disks) run separate software on separate hardware. This opens the door to server-embedded security that cannot be disabled by any software (even the OS) running on client systems as shown in the figure above. Of course, such servers have a narrow view of system activity, so they cannot distinguish legitimate users from clever impostors. But, from behind the thin storage interface, a self-securing storage server can actively look for suspicious behavior, retain an audit log of all storage requests, and prevent both destruction and undetectable tampering of stored data. The latter goals are achieved by retaining all versions of all data; instead of over-writing old data when a write command is issued, the storage server simply creates a new version and keeps both. Together with the audit log, the server-retained versions represent a complete history of system activity from the storage systems point of view.
Contact: Gregory Ganger is a Professor in the Electrical and Computer Engineering Department at Carnegie Mellon University.

Self-securing Storage and Self-securing Devices

Columbia University: Autonomizing Legacy Systems
Autonomic computing self-configuring, self-healing, self-optimizing applications, systems and networks is widely believed to be a promising solution to ever-increasing system complexity and the spiraling costs of human system management as systems scale to global proportions. Most results to date, however, suggest ways to architect new software constructed from the ground up as autonomic systems, whereas in the real world organizations continue to use stovepipe legacy systems and/or build systems of systems that draw from a gamut of new and legacy components involving disparate technologies from numerous vendors. The goal of this project is to retrofit

autonomic computing onto such systems, externally, without any need to understand or modify the code, and in many cases even when it is impossible to recompile. The project presents a meta-architecture implemented as active middleware infrastructure to explicitly add autonomic services via an attached feedback loop that provides continual monitoring and, as needed, reconfiguration and/or repair. The lightweight design and separation of concerns enables easy adoption of individual components, as well as the full infrastructure, for use with a large variety of legacy, new systems, and systems of systems.
Contact: Gail Kaiser is a Professor and the Director of the Programming Systems Lab in the Computer Science Department at Columbia University.

Autonomizing Legacy Systems

Cornell University: Astrolabe
The dramatic growth of computer networks creates both an opportunity and a daunting distributed computing problem for users seeking to build applications that can configure themselves and adapt as disruptions occur. The problem is that data often resides on large numbers of devices and evolves rapidly. Systems that collect data at a single location scale poorly and suffer from single-point failures. Astrolabe is a new system to automate self-configuration, monitoring, and to control adaptation. Astrolabe operates by creating a virtual system-wide hierarchical database, which evolves as the underlying information changes. Astrolabe is secure, robust under a wide range of failure and attack scenarios, and imposes low loads even under stress.
Contact: Kenneth Birman is a Professor in Computer Science at Cornell University working on Astrolabe.

Astrolabe

Georgia Institute of Technology: Qfabric
Distributed applications require end-to-end Quality of Service (QoS) management to ensure that (a) such applications achieve their goals in regard to functionality and performance and (b) system resources (processors, networks, disks, memory, etc.) are shared in a manner that prevents applications from interfering with each other. QoS-awareness of applications is an approach to allow them to take part in a resource management. This happens through interfaces that allow applications to specify their desired QoS or monitor the achieved QoS. Our approach is to closely integrate applications and resource managers in the QoS management. This is achieved by tying applications and resource managers through the same event-based control path. In other words, any control information exchanged between applications via the control path can be monitored by the underlying resource management. On the other hand, all resource management activities can be monitored by the application. Further, application and resource managers can interact freely to ensure optimal resource scheduling and adaptations.
Contact: Karsten Schwan is a Professor and the director of the Systems Research Group in the College of Computing at Georgia Institute of Technology.

Qfabric

IBM RESEARCH

Listed below are descriptions of and links to some of the projects

underway at IBM. Some of them have already appeared in products. Some are just beginning. This list of projects will grow as new areas of research emerge.

LEO -- DB2's Learning Optimizer
LEO is a comprehensive way to repair incorrect statistics and cardinality estimates of a query execution plan (QEP). By monitoring previously executed queries, LEO compares the optimizer's estimates with actuals at each step in a QEP and computes adjustments to cost estimates and statistics that may be used during future query optimizations.

In practice, LEO actually learns from its past mistakes, i.e., accelerating, sometimes drastically, future executions of similar queries while incurring a negligible monitoring overhead on query compilation and execution.

Storage Tank
Storage Tank is a new file system for Storage Area Networks that is being developed at Almaden Research. Major features of this system include heterogeneous file sharing, policy-based file and storage management, high performance, and scalability. This technology is currently used in Tivoli's Storage Manager.

Storage Tank Initiative promises universal access to data
IBM News release

Interconnected servers can balance the load
IBM Think Research Magazine

IBM plans to fund a regular stream of academic grants to support research in the area autonomic computing, and we call on others in the I/T industry to do the same. IBM will continue to flow money through our University Relations program to fund essential research like autonomic computing. For more specific information about such awards and fellowships please follow the links below.

- IBM Shared University Research (SUR) Program
- PH.D. Fellowships Program
- IBM Scholars Program
- Faculty Awards (Nominated by IBM Reseachers)

The IBM Ph.D. Fellowship Awards is an intensely competitive program which honors exceptional Ph.D. students in many academic disciplines and areas of study, for example: computer science and engineering, electrical and mechanical engineering , physical sciences (including chemistry, material sciences, and physics), mathematical sciences (including optimization), business sciences (including financial services, communication, and learning/knowledge), and service sciences, management, and engineering. Additionally, IBM pays special attention to an array of focus areas of interest to IBM and fundamental to innovation, including:

- **Analytics**
  Optimization, data visualization, business intelligence, particularly social networks & analysis, information-based medicine, services and software to improve business performance

- **Business transformation and services innovation**
  Labor based supply chains, SSME - Services Science, Management and Engineering
- **Expressive ways of manipulating computers**
  HCI, programming models and tools, language understanding, ontology's
- **Information management and analysis**
  Real-time data analysis, information based medicine, data base technology
- **Computing infrastructure**
  Management of computer centers virtualization, autonomic computing, power management
- **Pervasive computing**
  Sensors and actuators
- **High Performance Computing**
  Use of massive parallelism for non-scientific/engineering applications (including those applications), cell/multi-core, power management

The IBM Ph.D. Fellowship Awards program also supports our long-standing commitment to workforce diversity. IBM values diversity in the workplace and encourages nominations of women, minorities and all who contribute to that diversity.

Scope

The IBM Ph.D. Fellowships are awarded worldwide. IBM Ph.D. Fellows are awarded tuition, fees, and a stipend for one nine-month academic year. Stipends vary by country/geography and the student will be informed at the time of the award what the stipend is for their country/geography. All IBM Ph.D. Fellows are matched with an IBM Mentor according to their technical interests, and they are encouraged to intern at an IBM research or development laboratory under their Mentor's guidance. An IBM ThinkPad is awarded during the internship. Internship assignments are designed to strengthen and broaden the Awardee's technical experience and contacts. Interns are paid by their host site and will be subject to the prevailing terms and conditions of the internship program at that site. IBM is an equal opportunity employer.
IBM Ph.D. Fellowships are awarded for one nine-month academic year. After receiving an award, an Award Recipient may be renominated the following year for consideration to receive an IBM Ph.D. Fellowship again, based on the Award Recipient's continued exceptional academic standing, progress and achievement, and sustained interaction with IBM's technical community. A student may compete annually and be awarded a maximum of three years. IBM requests that a maximum of two nominations per department be submitted in addition to any renewal nominations.

Eligibility

Students must be nominated by a faculty member. They must be enrolled full-time in a college or university Ph.D. program, and they must have completed at least one year of study in their doctoral program at the time of their nomination. Award Recipients will be selected based on their overall potential for research excellence, the degree to which their technical interests align with those of IBM, and their academic progress to-date, as evidenced by publications and endorsements from their faculty advisor and department head. While students may accept other supplemental fellowships, to be eligible for the IBM Ph.D. Fellowship Award they may not accept a major fellowship in addition to the IBM Ph.D. Fellowship.
For further information, see frequently asked questions or contact phdfellow@us.ibm.com.