iSeries V5R2 and IBM's autonomic computing initiative

Self-Managing IT Infrastructure
Technologies for e-business
Today's Diverse e-business Infrastructure

- Customers
- Business Partners
- Suppliers
- Employees

Middleware

- Directory and Security Servers
- Web Presentation Servers
- Web Application Servers
- Transaction Servers
- Data Servers

Edge Servers

Storage
Notes: Today's Diverse e-business Infrastructure

Managing heterogeneous systems is a fundamental challenge of e-business that exists today and the challenges will only grow. It is a diverse world from a server and workload perspective. Different market segments require a variety of architectures, platforms and technologies. Today, customers on average have a minimum of five different architectures to meet their IT computing needs. This demand can only be addressed today by the four eServer architectures core value propositions. Within five years, it is expected that hundreds of millions of people will be connected via wireless and other devices to the Web, driving trillions of transactions and tremendous amounts of rich media like voice and video accessing terabytes of storage/data. The supporting infrastructure will be very diverse, including appliances, far-flung branch servers, mainframes, and systems hosted by service providers. Compounding this problem, according to one IT analyst, the world will be short at least a million IT administrators by then, assuming the use of today’s technology.
Current eServer Product Line

**xSeries**
- Affordable, Intel-based servers with mainframe-inspired reliability technologies

**iSeries**
- Integrated, easy-to-use, e-business mainframes for the mid-market

**pSeries**
- Most powerful, technologically advanced UNIX servers

**zSeries**
- First enterprise servers designed from the ground up for e-business

Linux
Needed: Horizontal Integration Across Systems

- Systems Management and Automation
- Workload and Performance Management
- Security
- Availability and Serviceability Management
- Logical Process and Resource Management
- Clustering Services
- Database
- Connectivity Management
- Physical Resource Management

Supported Operating Systems:
- z/OS
- OS/400
- AIX
- Linux
- Win
- Solaris
Managing the Infrastructure

Workloads

New Workloads

Skills Shortage

Services and Software Costs

Complexity
Notes: Managing the Infrastructure

In an IT environment where workloads are growing, both in transaction volume and in complexity and where finding the right skills to manage servers and systems, the need for tools to manage the processes an data on which the business processes rely, automation becomes a necessity. Implementing and managing these tools will over time become a similar challenge, which can be addressed only by making these tools integrated into the service delivery tools and to deliver them as part of the IT infrastructure.

The horizontal integration of applications and system function provides the basis for integrated management tools, which will require little or no setup to perform the base management function to support the underlying functions.
IBM's autonomic computing initiative and the iSeries

OS/400 exploits IBM's blueprint for self-managing systems

- Self-protecting
  - Enterprise Identity Mapping
  - Digital signatures and intrusion detection

- Self-optimizing
  - Dynamic LPAR
  - Self-learning DB2 UDB Automatic Index Advisor

- Self-healing
  - Switched disk support for improved availability
  - Agent Building Learning Environment (ABLE) for problem management*

- Self-configuring graphical wizards for managing
  - Performance and multiple workloads
  - Switched disk clusters and high availability
  - Security and network
  - Storage and system backups

*NOTE: ABLE, an IBM Research Project, is designed to reduce human intervention through using artificial intelligence when diagnosing hardware and software errors. ABLE will debut first with iSeries and OS/400 V5R2. See: http://www.alphaworks.ibm.com/tech/able

Notes: IBM's autonomic computing initiative and the iSeries

Many of the founding elements for IBM's autonomic computing initiative and the iSeries already manifest into today's iSeries systems. One year after announcing the IBM's autonomic computing initiative to develop self-managing computer environments, iSeries continues to exploit IBM's blueprint for delivering technology and tools to ease management of systems.

V5R2 continues to build on many of the autonomic computing initiative elements that were available with V5R1 such as, such as self-optimizing dynamic LPAR and workload management, self-configuring graphical wizards, self-healing performance monitors, and self-protecting digital certificates. Extensive additional graphical wizards have been added with V5R2 to the iSeries Navigator to automate several complex configuration tasks along with increased flexibility to monitor and manage storage, jobs, and database tasks.

Some of the highlights for V5R2 include:

- Self-protecting Enterprise Identity Mapping for easing user identity management
- Self-protecting Digital certificate APIs for ISV applications to assist with unauthorized application modifications
- Self-optimizing dynamic LPAR to allow resource movement for virtual processing units between Linux and OS/400 partitions
- Self-optimizing index advisor and statistic collections for DB2 UDB for OS/400, allowing users to avoid manual tasks associated with SQL optimization
- Self-healing Independent disk pools for switched disk clustering
- Self-healing Agent Building Learning Environment (ABLE) enablement through building intelligent agents on the iSeries to assist with problem determination and diagnosis processes. The ABLE research project is made available by the IBM T. J. Watson Research Center. For additional information, see: [http://www.alphaworks.ibm.com/tech/able](http://www.alphaworks.ibm.com/tech/able)
- Extensive self-configuring graphical wizards to simplify network, performance, security, storage, work management and LPAR configuration tasks.

Note 1: Autonomic Computing reflects a vision to develop and deploy intelligent systems that self manage and regulate themselves, much the way the human autonomic nervous system manages the human body. This vision is motivated by the tremendous complexity in today’s computing environments and the resultant difficulties, and expense, of managing them. The biological metaphor suggest a systemic approach, coordinating activity across the many components of computing systems, achieving a much higher level of automation. For a complete discussion of the autonomic computing direction see the Autonomic Computing Manifesto ([http://www.research.ibm.com/autonomic/manifesto/](http://www.research.ibm.com/autonomic/manifesto/)).

Note 2: The graphic shown represents the four groupings of self managing activities according to IBM's autonomic computing initiative. You may also hear this graphic referred to as CHOP circles - self-Configuring, self-Healing, self-Optimizing, self-Protecting.
Notes: IBM's autonomic computing and the iSeries -2

Self-optimizing provides functions for servers to efficiently maximize resource utilizations with little to no human intervention. Across the entire eServer family functions are available today within each individual series to self-optimize.

Self-configuring is the ability for servers to define themselves "on-the fly". The most notable function available across the entire eServer family that is available today is Capacity Upgrade on Demand. However, each series has additional features and functions that are also available today both at the hardware level as well as operating systems level.

Self-protecting enables servers to avoid intrusion and provide security in an e-business environment. The e-Server platforms have been focused on the Internet environment, and this has driven focus in certain areas (example hardware encryption to drive SSL performance).

Self healing enables a server to recover from a failing component by taking the affected component off-line, fixing the failure, and/or finally reintroducing the component into service without any application disruption. Across the entire eServer line there are many functions that are available within all series today. These include: "N+1" Power which is redundant power that prevents an outage due to power failure, "N+1" Cooling which is redundant cooling that prevents an outage due to a cooling failure, Dual Power Cords which enable dual source power to the server, Hot plug Power which allows the replacement of failing power supplies during normal system operation, Hot plug cooling which allows the replacement of failing fans/chillers during normal system operation, ECC Memory which is Error Correcting Code on main memory to fix single bit and detect double bit soft or hard failures, ECC Caches which is Error Correcting Code on cache memory to fix single bit and detect double bit soft or hard failures, Cluster support to provide system redundancy for availability and scalability, Memory Chipkill (TM) automatically and transparently removes a failing dimm from the configuration and substitutes a spare dimm in its place, Service Processors which monitor the system, performs diagnostic tests, records errors, Call home to request service and support remote service, and RAID Storage which is disk storage available in RAID format that allows for individual disk failures to be replaced during normal operations without losing access to data.
Self-configuring
Self-configuring is the ability for servers to define themselves "on-the fly". The most notable function available across the entire eServer family that is available today is Capacity Upgrade on Demand. However, each series has additional features and functions that are also available today both at the hardware level as well as operating systems level.

iSeries Self-Configuring features:
- Integrated xSeries Server allows Windows Server running under iSeries, with iSeries storage management helps manage the PC Servers in the IT environment.
- Workload Balancing/Job Prioritization, queue priority, time slice, working set size
- EZ-Setup wizard for quick system set up and customization
- Automatic detection/configuration of hardware resources/devices
- Autostart and prestart job entries
- Automatically starting TCP/IP interfaces and servers when TCP/IP is started
- Automatic configuration of communication configuration objects (APPC/APPN, remote WS, Telnet, pass-through, etc.)
- Disk Balancing based on capacity and data access
- Automatic object placement and additional storage allocation
Self healing
Self healing enables a server to recover from a failing component by taking the affected component off-line, fixing the failure, and/or finally reintroducing the component into service without any application disruption. Across the entire eServer line there are many functions that are available within all series today. These include: "N+1" Power which is redundant power that prevents an outage due to power failure, "N+1" Cooling which is redundant cooling that prevents an outage due to a cooling failure, Dual Power Cords which enable dual source power to the server, Hot plug Power which allows the replacement of failing power supplies during normal system operation, Hot plug cooling which allows the replacement of failing fans/chillers during normal system operation, ECC Memory which is Error Correcting Code on main memory to fix single bit and detect double bit soft or hard failures, ECC Caches which is Error Correcting Code on cache memory to fix single bit and detect double bit soft or hard failures, Cluster support to provide system redundancy for availability and scalability, Memory Chipkill TM automatically and transparently removes a failing dimm from the configuration and substitutes a spare dimm in its place, Service Processors which monitor the system, performs diagnostic tests, records errors, Call home to request service and support remote service, and RAID Storage which is disk storage available in RAID format that allows for individual iSeries features:

- Memory Scrubbing which continuously scan memory in the background for errors during normal operations
- ECC L1 D cache: Error Correcting Code on the L1 Data Cache is provide to allow for transparent single bit error correction and double bit error detection.
- Dynamic CPU/Memory deallocation which takes CPUs or memory off-line when errors are detected and unrecoverable allowing the remaining resources to continue.
- Persistent CPU deallocation keeps the CPU off-line even after an IML.
- Hot swap storage allows disks maintenance without impact in normal operations
- Auto reboot recovery will automatically reboot the system following an unrecoverable software error, software hang, hardware failure or environmentally-induced (AC power) failure
- Service Processor Automatic Restart provides a means to restart a service processor upon failure in an attempt to clear the error condition
- Hot swap PCI (storage) allows disks to be maintained without impact in normal operations
- Redundant I/O buses which is loop technology that allows for redundant paths to I/O drawers. Redundant I/O availability features include CRC checking on the bus with packet retry on bus time-outs. In addition, if a link fails, the hardware will automatically initiate a bus reassignment to route the data through the alternate path to its intended destination.
- PCI Enhanced Error Handling which is customized technology to isolate and recover from PCI errors
- Predictive Thresholds on processor, cache, memory, and RIO link recovered errors
- Dynamic IP takeover in a clustered environment allows for a clustered system to automatically take over work from a failed system, in a manner transparent to the end user.
Notes: IBM's autonomic computing and the iSeries -5

Self healing continued

- IP Load Balancing with Virtual IP and Route Fault Tolerance, allows for better link utilization for IP traffic, as well as redundant paths for availability.
- APPN High Performance Routing - ability to automatically detect and reroute around failures within the network
- Automatic switchover to another n-way processor if one fails during a self-test during system IPL to complete IPL even in a degraded mode
- Automatic re-IPL and mainstore dump on a software crash (options to dump contents to disk automatically can be optionally set) allows for quicker recovery times in event of a system crash.
- Automatic I/O reset/reload allows for self healing of specific I/O resources that are critical to the operation of the system.
- Software-managed thresholding of soft errors allows an operator to predict possible hard errors, and provide fixes before an outage occurs.
- Unattended save and restores improves the efficiency of these operations, and reduces operational costs.
- Communications error recovery - detect and automatically recover from intermittent failures.
Notes: IBM's autonomic computing and the iSeries -6

Self-optimizing
Self-optimizing provides functions for servers to efficiently maximize resource utilizations with little to no human intervention.

iSeries features:
- Monitoring, reporting, and the ability to take actions - including to wireless devices. Monitoring (through Management Central a GUI based interface) allows thresholds to be set and trigger an application action - such holding a job or informing a mobile pager. Monitor types supported include:
  - System (over 30 performance metrics, including CPU utilization, disk I/Os, database capability)
  - Job monitors (includes job performance metrics and job status changes)
  - Message (includes types of messages or specific message id sent to a message queue and, messages sent to a job log)
  - File monitor (includes files/tables and the OS/400 History Log) for changes to the file/table)
  - B2B transactions monitors for transactions originated through Ariba or other markets. This allows detailed transaction status and management.
- Performance Advisor provides a means to use monitored performance metrics, and propose tuning options to increase overall performance.
- Automatic enablement of expert cache algorithms allows tuning based on heuristic information.
- Capacity planning, and future resource forecasting allows customers and sales personnel to size upgrades or new systems based on current performance data and new workload estimations. (aka PM400)
- Dynamic priority scheduling allows for system priorities to be adjusted without system or job restarts.
- Self-optimizing provides functions for servers to efficiently maximize resource utilizations with little to no human intervention. Across the entire eServer family functions are available today within each individual series to self-optimize:
  - Job wait accounting is a way to measure current workloads and specifically identify where resources are being used.
  - Built in Job Scheduler allows unattended task kick-off via a predetermined events with spool management allows you to graphically manage output files.
  - Guaranteed Network Quality of Service on IP communications allows for the best service delivery to end users. Extending this for all resources with Work Load Management functionality is the planned direction
  - Automatic Data Spreading on disk allows for improved performance and data management, based on the single level store architecture.
  - Hierarchical storage management (dynamic archival and retrieval of objects) allows for the data management that is set up by the administrator, but basically transparent to the user.
  - Automatic responses to messages (system reply list) both informational and operational. This allows the operator to preselect how specific operations should be performed, and then have the system automatically take actions.
  - PRSM hardware capabilities to optimize hardware resources.
Self-protecting
Self-protecting enables servers to avoid intrusion and provide security in an eBusiness environment. The e-Server platforms have been focused on the Internet environment, and this has driven focus in certain areas (example hardware encryption to drive SSL performance).

All the e-Servers have common self-protecting elements which are amiable today and are listed as follows:

- **LDAP**
  Lightweight Directory Access Protocol which provides an Open Standard directory support, that allows the directory to be queried from the current server or any system server within the enterprise that supports LDAP. Many of the deployed LDAP directory services are being used to store Security Related information. Directory servers provide a granular access control model to secure information stored in the directory. Directory servers utilize many of the security features in eServer systems such as SSL, Kerberos, and hardware cryptographic support as well as features unique to the security services offered on the eServer platforms.

- **Kerberos**
  Kerberos is not a new authentication method, but as Microsoft chose to add this support to their operating system, Kerberos support was added to all the eServer platforms. This allows us to interoperate with Windows and all the other eServer platforms. For example, if you logon to Windows 2000, the user can gain a Ticket Granting Ticket (TGT) to be passed to another Server, and that TGT can be accepted by an eServer server.

- **SSL**
  Secure Sockets Layer is the defacto standard of how people encrypt sensitive data on the internet. SSL drives encryption at the application level, where VPNs create tunnels where all the data is encrypted between two points. iSeries and zSeries provide general purpose interfaces to customers to create their own SSL, and all the e-Servers incorporate SSL code within their Operating Systems. AIX provides JSSE and SSLite for Java applications. When a digital certificate is used for SSL, the Server/Application need to know it is a valid certificate (just like credit cards, where they check if it is revoked, or stolen). LDAP is used to determine the validity of the certificate that represents the remote entity.

- **VPN**
  Virtual Private Networks are a way to create a secure tunnel between two places, where all the data flowing is encrypted. This takes the burden off of the application trying to decide what is sensitive, and encrypts everything. All the eServers have participated in "VPN Bakeoffs", where all the leading VPN vendors come and do interoperation testing. On z-Series, VPN configurations can be stored in a directory service. The Interoperability of the eServers with other platforms has been verified by successful participation and certification by the ICSA Labs IPSec Product Program.
Self-protecting continued

- **Digital Certificates**
  Digital Certificates are more secure than Userid's and Passwords. For this reason, more applications are moving to utilize them. iSeries and zSeries have system services to create digital certificate and the encryption key pairs that are needed, and also correlate them to the local identity (RACF or OS/400). AIX supports digital certificate use in their VPN/IPSec solution. The leader in Digital Certificates is Entrust. LDAP in z-Series has been Entrust Certified to be used with their product, which allows the Certificate Revocation Lists (CRLs) to be stored on z/OS LDAP.

- **Hardware Encryption**
  For years, IBM has had hardware encryption products, but over the past several years, the 4758 product was created for the use of all eServer platforms (x, i, p, and z) to help 1) the financial and e-business companies have a highly secure encryption product, that would adhere to strict government laws, and 2) provide a way to speed the SSL handshake freeing up the processor to do other work. Since data is not safe on the internet, unless it is encrypted (usually using VPNs or SSL), being able to do it quickly, and securely is a critical aspect of our Servers.

- **GSKIT**
  GSKit is a common toolkit that is reused within 60+ products across IBM. It implements the SSL protocol.
Trademarks and Disclaimers

8 IBM Corporation 1994-2002. All rights reserved.
References in this document to IBM products or services do not imply that IBM intends to make them available in every country.
The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

<table>
<thead>
<tr>
<th>Product</th>
<th>Description</th>
<th>Trademark or Registered Trademark</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRMS</td>
<td>Host Integration Series</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>ADSTAR</td>
<td>Host on Demand</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>Advanced Function Printing</td>
<td>ClusterProven</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AFP</td>
<td>HTTP Server for AS/400</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AIX</td>
<td>Net.Data</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AnyNet</td>
<td>IBM Logo</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>Application Development</td>
<td>DB2 Extenders</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>APPN</td>
<td>DB2 UDB for AS/400</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AS/400</td>
<td>DB2 Universal</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AS/400e</td>
<td>e-business logo</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>AT</td>
<td>e(logo) Server</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td>BrioQuery</td>
<td>Enterprise Storage Server</td>
<td>IBM Corp.</td>
</tr>
<tr>
<td></td>
<td>iSeries</td>
<td>IBM Corp.</td>
</tr>
</tbody>
</table>

cc:Mail, Domino.Doc, Freelance, LearningSpace, Lotus, Lotus Domino, Lotus Notes, iNotes, QuickPlace, Sametime, and Word Pro are trademarks of Lotus Development Corporation in the United States, other countries, or both.

Tivoli and NetView are trademarks of Tivoli Systems Inc. in the United States, other countries, or both.

C-bus is a trademark of Corollary, Inc. in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

PC Direct is a trademark of Ziff Communications Company in the United States, other countries, or both and is used by IBM Corporation under license.

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

IBM’s VisualAge products and services are not associated with or sponsored by Visual Edge Software, Ltd.

Linux is a registered trademark of Linus Torvalds.

UNIX is a registered trademark of The Open Group in the United States and other countries.

SET and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

Other company, product and service names may be trademarks or service marks of others.

Information is provided “AS IS” without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information in this presentation concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

Some information in this presentation addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM’s current investment and development activities as a good faith effort to help with our customers’ future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.