e-business on demand
a technology perspective

Open Standards
- Web Services
- Autonomic Computing
- Grid Computing
- e-Utility

Fulvio Capogrosso
Distinguished Engineer
Server Group, South Region, EMEA
Agenda

- Scenario
- Definitions
  - business proposition
  - technical proposition
  - financial proposition
- The Roadmap
- Key Enabling Technologies
  - autonomic computing
  - grid computing
  - integration technologies

Open Standards

- Web Services
- Autonomic Computing
- Grid Computing
- e-Utility
Today's e-business Scenario

Unpredictable Demand

Business Efficiency

Time to Market

Operational Flexibility

Middleware

Directory & Security Servers

Web Application Servers

Transaction Servers

Data Servers

Edge Servers

Customers

Suppliers

Partners

Employees

Storage

Web Presentation Servers

Web Servers

Transaction Servers

Difficult Economy Cycle

Business Continuity

Cost & ROI (Return-on-Investment)

Complexity Management

Cost & Time to Market

Employees & Suppliers

Partners & Customers
Technology Evolution

Internet & Connectivity

Base Technology

Grid Computing

Autonomic Computing
Open Standards
The foundation for operational flexibility

Linux

WSDL

SOAP

XML

Java

Web Services

OGSA
e-business on demand Summary

- 'e-business on demand' is the term IBM has chosen to name a new movement in business and computing, which is changing the way technology is deployed and used in business.

- An 'on-demand' Enterprise is defined by a set of:
  - Business attributes
  - Key characteristics of the operating environment
  - New environment economics

### Business attributes
- Responsive
- Variable cost
- Focused
- Resilient

### IT Attributes
- Open
- Integrated
- Virtualized
- Autonomic

### New Economics
- Computing as a utility
- Variable costs
- Variable ownership
e-business on demand

Suppliers
- Application Providers
- Service Providers
- Network Providers
- Technology Providers

Enterprise

Customers

Responsive

Focused

Resilient

Variable
Sense and Respond Business Model

An adaptive management framework developed at the IBM Advanced Business Institute

- Business on demand rather than business as planned
- Business strategy, structure and governance principles designed for environments of unpredictable changes
- New business core competencies
  - earlier understanding of what is happening now and in the near future
  - dynamically dispatch modular capabilities to respond
  - operational and management flexibility
The Path to 'on demand' IT

**Business Process Optimization**
- Business process integration
- Business process re-ingenineering

**IT Infrastructure Optimization**
- including variable cost commercial offerings and e-Utility services

**Higher Levels of Virtualization**
- Grid computing
New 'on-demand' Economics

More flexible, variable and economically attractive choices for buying and managing computing

- Data center virtualization to build 'internal computing utilities'
  - higher resources utilization
  - open standards
  - permanent and On/Off Capacity on Demand
- External computing utility
  - to acquire computing and applications from service providers
- 'pay per use'
**Key Enabling Technologies**

**Autonomic Computing**

**Grid Computing**

"Simple .. Open .. Standard .. Dynamic .. Virtual"

[Diagram showing self-configuring, self-healing, self-optimizing, and self-protecting features]

*Web Services*
Is the term IBM uses for systems that have the ability to manage themselves and dynamically adapt to change in accordance with business policies and objectives.
Autonomic Computing

Autonomic Computing Reference Architecture

- 4 major features
  - Self-configuring
    - Adapt automatically to dynamically changing environments
  - Self-healing
    - Autonomic problem determination and resolution
  - Self-protecting
    - Systems designed to protect themselves from unauthorized access
  - Self-optimizing
    - Monitor and tune resources automatically

- Delivered through multiple R&D projects
  - Open standard based
  - Leveraging on high level features already existing in current high end server technology

- 5 levels of implementation maturity
  - Basic
  - Managed
  - Predictive
  - Adaptive
  - Autonomic
Autonomic Computing Delivery

- Core Enabling Capabilities
  - Problem Determination
  - Common System Administration
  - Autonomic Monitoring
  - Solution Install
  - Policy-based Management
  - Complex Analysis
  - Enterprise Workload Management

- Systems & Resources Management Technologies
  - Enterprise Workload Management
  - Project Symphony
The eWLM Environment

- Multiple heterogeneous servers
- Complex transaction flow
- Dynamic transaction topology

Quality of service

Income txns
Enterprise Workload Manager

- System resources management software for end-to-end autonomic performance management of transaction workload executing on heterogeneous multi-tiered configurations

- Key elements:
  - policy based end-to-end active management to business priorities
  - customer defined performance goals for business importance classes
  - dynamically discovery of workload topology
  - self-learning tuning algorithms from accumulated environment knowledge

- Distributed system composed of two parts: a global eWLM manager and multiple instances of local eWLM managers
  - Java application running on any Java supporting platform
  - local eWLM for AIX, Solaris, Windows, zOS, OS/400, Linux, (HP-UX)

- The global eWLM manager will provide knowledge based services to external users (ex. other system management applications)
**eWLM management domain**
- Set of servers (OS instances) under the eWLM control

**eWLM Management policy**
- Administrative definition of performance expectations for the domain
- Concrete definition of the difference between good and bad performance

**eWLM Managed servers**
- Local platform eWLM (LeWLM)
- Control point for local activities

**eWLM management server**
- Global eWLM (GeWLM)
- Control point for global activities
**eWLM Algorithms**

**eWLM Policy**

- Learn server role in end-to-end flows
- Monitor resource consumption & delays
- Identify service classes needing attention (e2e)
- Identify potential resource adjustments
- Project impact of adjustments
- Initiate resource adjustment(s)

**Local eWLM**

- When local resources adjustments are not sufficient, to meet the end-to-end goals, the global eWLM makes further adjustments

**Locally managed resources**
- system resources (CPU, Mem, I/O)
- network resources (bandwidth,....)
- LPAR resources (physical, logical processors,..)
- VMWare resources (partitions,..)

**Globally managed resources**
- application traffic (load balancers)
- server instances (provisioners)
eWLM Platform Instrumentation

- **Middleware level**
  - Extensions to The Open Group ARM Standard, V3
    - Process registration, deregistration
    - Work request classification, start, and stop
    - Work request execution thread bind, unbind
  - Extensions accepted and expected to be approved by 1Q2003

- **Operating system level**
  - Resource management & optimization layer
each work request has a unique 'transaction correlator'
application flows correlator along with each external request
arm_start_transaction API accepts an optional parent correlator
correlator propagates service class, topology, reporting relationship
**eWLM Management Server Data**

### Local eWLM

- **Application Environments**
  - Application Identity
  - Process Relationships
- **Service Classes**
  - Policy Objective
  - End-to-End Results
- **Segment Resource Data**
  - Resource Consumption
  - Resource Delays

**Application participation in end-to-end flow**

### Global eWLM

- **Application Environments**
  - Application Identity
  - Summary Statistics
- **Service Classes**
  - Policy Objective
  - End-to-End Results
- **Participating Servers**
  - Server Identity
  - State information
  - Summary statistics

**Relational Topology**
The IBM Software Approach

- Systems Management & AD Environment
  - Applications
  - Application Software
  - Operating System(s)
  - Platforms
  - Hardware and Architecture(s)

- Systems Management
  - Rational
  - WebSphere

- AD Environment
  - CRM
  - ERP
  - PLM
  - SCM

- WebSphere

- DB2

- Lotus

- IBM and non-IBM Op. Systems, Platforms
  - IBM eServer and non IBM hardware
IBM Tivoli Software in the Solution Lifecycle

- Capture Requirements
- Model and Design
- Analyze and Report
- Develop & Test
- Monitor and Manage
- Deployment and Configuration
- Correlation and Automation
- Business Impact Management

Run Time Environment:
- WebSphere
- DB2
- Lotus
- ISVs
An Orchestrated Environment

Sense
- detect an increase in demand and deterioration in performance

Monitor
- automatically

Knowledge

Plan

Execute

Respond
- reallocate existing resources according to business requirements (provisioning)

Middleware
- servers
- operating systems
- middleware
- applications
- network devices

Storage

Edge Servers

Web Pres. Servers

Web Application Servers

Transaction Servers

Directory & Security Servers

Database Servers

Income txns
**Traditional Approach**

- Individual dedicated execution environments sized for estimated peak capacity
  - over-provisioning
  - low overall utilization
  - static allocations versus dynamic changing environments
Orchestrated Provisioning

Service Level Management

Capacity Planning

Infrastructure Provisioning

resource pool

monitoring
Project Symphony

a family of modular on demand technologies and solutions for infrastructure resources orchestration with no human intervention

- IBM Tivoli Intelligent (Thinkdynamics) Orchestrator
  - to manipulate the IT environment in real time according to pre-defined business policies
  - includes automatic provisioning of IT resources

- IBM Tivoli Provisioning Manager
  - part of the Orchestrator solution
  - automates the provisioning process of infrastructure deployment
  - framework to collect and execute best practices for datacenter management

- IBM Web Infrastructure Orchestration
  - pre-tested, pre-integrated solution for web serving environments
  - software, hardware, services
  - includes pre-packaged intelligence
IBM Web Infrastructure Orchestration

- **What it does:**
  - Intelligently manages changes in web demands by automatically repurposing new blade servers to meet service level agreements

- **Capability:**
  - Pre-tested, pre-integrated orchestration of web servers on Blades with optional implementation services

- **Key components of the “soft bundle” solution:**
  - IBM Tivoli Intelligent ThinkDynamic Orchestrator
  - Orchestrator workflows and IBM Director scripts
  - IBM WebSphere
  - IBM DB2
  - IBM eServer BladeCenter
  - IBM FASST900 Storage Server
Grid computing is a distributed computing model which supports the concept of virtual dynamic organizations by providing secure and coordinated access and sharing of heterogeneous and geographically distributed resources (applications, data, processing power, storage capacity and other) over a network and across organizational boundaries, using a set of open standards and protocols.
Resource Virtualization

*a definition*

The effect of applying hardware and/or software technology to a set of physical (real) resources to make them appear to the user as a set of (virtual) resources with different (better) characteristics.

Ex. processing servers virtualization

Other virtualization technologies:
- virtual file systems
- virtual disk storage (storage tank)
- virtual database (federated database)
DB2 Family
Sybase
Informix
SQL Server
Oracle
Teradata

Wide variety of Clients
Single virtual database view

Single point-of-connect” for the end-users and applications
The Globus Project

- An initiative by the Argonne National Laboratory, the University of Southern California and the University of Chicago (1998) with the objective to develop fundamental technologies needed to build (computational) grids.
  - many collaborators including IBM and Microsoft
  - many sponsors including government agencies, IBM, Microsoft, Cisco
  - affiliated with many WW organizations such as the Global Grid Forum (GGF)

- The Globus Project provides the Globus Toolkit, a collection of basic grid services and APIs that can be used to build computational grids and grid based applications (open source (Globus Toolkit Public License (GTPL))
  - Globus Toolkit GT2.4
  - IBM Grid Toolbox for GT2.0 for AIX and Linux

- Open Grid Services Architecture (OGSA)
  - result of an R&D 2001 program with IBM participation
  - delivered by Globus as Globus Toolkit 3.0 (GT3)
  - GRID Services specifications (as extensions of WEB Services)
  - to be delivered with ALL IBM eServer platforms
OGSA & Quality Of Service Delivery

- QoS is traditionally delivered at platform level through vertical integration of multiple technologies
  - "stovepipe solution towers"
  - local platform dependent interfaces

- At the IT Infrastructure level (multiple heterogeneous platforms) the end-to-end QoS delivery is fractured
  - labor intensive re-integration of QoS delivery
  - across multiple platform
  - across multiple implementations
  - across multiple standards
Open Grid Services Architecture (OGSA)

Distributed Resource Management Across Heterogeneous Platforms

- Meta-Operating System functions for seamless QoS delivery
  - standard service definitions independent from the local (platform) implementation
  - GRID Services

- Common Base for Autonomic Management Solutions
  - autonomic management interfaces exposed as GRID services

- Supports the creation of virtual resources as stateful OGSA service instances
OGSA Based Architecture Framework

Common autonomic construct for all system elements
Distributed components and systems integrated as one virtual operating system
Grid Computing

To enable secure and coordinated access and sharing of heterogeneous and geographically distributed resources (applications, data, processing power, storage capacity and other) across organizational boundaries.
Grid Computing

Open Grid Services Architecture (OGSA)

Grid Middleware (OGSA)

Physical resources

Virtual environments

Users
A WebSphere on demand Technology

- SLAs protection by coordinating use of available resources
  - local grid, remote grid, provisioning
- Multiple mixed workloads on the same GRID infrastructure
- Reactive mode (monitoring) or predictive mode (active forecasting)
Resource Virtualization can help to optimize the IT infrastructure and:

- Reduce overall complexity
- Standardize policies, procedures and operations
- Drive adoption of open standards, end-to-end
- Simplify systems management
- Facilitate business continuity and disaster recovery
- Enhance resources utilization and optimize scalability
- Optimize performance, availability, operational resiliency
- Improve total cost of ownership (TCO)
- Accelerate e-business integration
- Position IT infrastructure for the on-demand business model
Why Linux

- **IT Infrastructure Optimization: Linux is a key player**
  - Linux lowers Total Cost of Ownership (TCO)
  - Linux allows application portability
  - Linux enables higher levels of manageability (ex. backup and recovery)
  - Linux provides operational flexibility (ex. test and temporary environments)

- **CIO/CEO talking points**
  - Linux is an Open Source operating system
  - Linux can be tailored to meet individual business needs
  - Linux is free or available at very low cost
  - Linux runs on more hardware platforms than any operating system available today
IBM Commitment to Linux

"IBM has made all of its server platforms Linux friendly, including zSeries, iSeries, pSeries and xSeries servers"

- IBM commitment:
  - has invested $1bn USD+ in Linux related development
  - has invested additional $300mill USD into Linux services development (implementation, consulting, education, ...)
  - has created a new organization for Linux
  - has made all IBM server platforms Linux-ready (iSeries, pSeries, xSeries & zSeries)
  - has engaged closely with the Linux community to help Linux evolve (7,000 IBMers working in porting, research, services centers and development labs)
  - has made IBM technologies available to Linux and open source communities
  - has initiated Joint Customer Project for business and technical validation
  - has announced the IBM products Linux ready & established Linux distributions (more than 50 software products are available on Linux - Data Management, WebSphere Software, Lotus and Tivoli Software)
  - Over 4,200 ISV applications have been ported to Linux
  - has disclosed plan for next 12 months......
  - has worked on more than 6,500 customer Linux projects in 2002
  - 120,000 competitive servers have been replaced by IBM Linux Servers
Responsive
Variable
Focused
Resilient

Business Objectives
Initiatives
IT Attributes
Value

Business Transformation
Operating Env.
Utility Services
Open
Integrated
Virtualized
Autonomic

Business Efficiency
Operation Flexibility
Time to Market
Competitive Advantage

Value
End Presentation

Fulvio Capogrosso
Distinguished Engineer
Server Group, South Region, EMEA