Architectural Approaches to Cloud Computing

Michael Jackson
President
Adaptive Computing
Virtualizing the Data Center: From Silos to Clouds

Automation and service centricity will take that pool of resources and do useful work based on business policies and service-level requirements. Also, the decoupling created by virtualization combined with defined service offerings and automation are great enablers of cloud computing. Essentially, real-time infrastructure (RTI) is a private cloud-computing engine.

The services-oriented interfaces created between an RTI and its customers enable an IT organization to leverage external cloud services where appropriate. Cloud computing is the inevitable result of a service provider evolving to RTI.
“Workload mobility is perhaps the most important step toward a dynamic data center. The data center's dynamism opens up a world where the IT administrator or automation software can move workloads to achieve a strategic IT goal. All other aspects of the dynamic data center are predicated on this ability.”
Alternate Approaches

**VM-Based (VM):**
- VM: Miss 75% of Workloads
- VM: Lock-in to single hypervisor
- VM: VMWare is expensive path
- VM: Limited organizational-level management (Obj., Org., etc.)

**Workflow/Script-Based (VM):**
- W: Takes hard-coded paths
- W: Define every scenario
- W: Breaks with complexity
- W: Static
- W: Requires human administration

**Intelligence Engine-Based (IE):**
- IE: Virtual and Physical
- IE: Multi-hypervisor support
- IE: Use most cost effective options
- IE: Rich organizational-level mgmt

**Intelligence Engine-Based (IE):**
- IE: Makes best decision
- IE: Define principles
- IE: Scales to complexity
- IE: Dynamic
- IE: Able to be Fully Automated
Alternate Approaches

Peer RM/Fabric-Based (P):
- P: Lacks Unified View/Decision
- P: Embedded to middleware
- P: More Homogeneous Infr.
- P: Filters, less efficient actions
- P: Lock-in to more of expensive resource management layer

Mechanism-Based (M):
- M: Acts, but not intelligently
- M: Self-service, but not WL-driven
- M: Limited SLA management
- M: Very limited optimization
- M: Needs administrator
- M: Very limited conflict resolution

Intelligence Engine-Based (IE):
- IE: Unified Automation Decisions
- IE: Abstract, multi-sourcing
- IE: More Heterogeneous Infr.
- IE: Optimizer
- IE: Use best-fit and commodity resource managers

Intelligence Engine-Based (IE):
- IE: Makes best decision
- IE: Self-service and WL-driven
- IE: Rich SLA management
- IE: Optimized
- IE: Fully automates
- IE: Rich conflict resolution
Alternate Approaches

Cloud Onramp-Based (O):
- O: Move, but not manage
- O: Lacks intelligent escalation
- O: Lacks SLA management
- O: Does not manage local resources
- O: Lacks best selection of workloads to migrate

Web Services/Load Balancer-Based (L):
- L: Limited class of WL (Java, .net, etc.)
- L: Limited SLA management
- L: Limited to no escalation
- L: No self-service project space

Intelligence Engine-Based (IE):
- IE: Moves and manages life cycle
- IE: Rich escalation
- IE: Rich SLA management
- IE: Manages remote and locally
- IE: Use best-fit workloads and resources

Intelligence Engine-Based (IE):
- IE: Open set of workloads
- IE: Rich SLA management
- IE: Rich escalation
- IE: Workload-driven and self-service project spaces
Unified Automation Intelligence

- Open Heterogeneous Cloud
- Avoid Expensive Lock-In (at Hypervisor, Hardware, or middleware layer)
- Rich SLA Management
- Rich Escalation and Conflict Resolution
- Resource Sharing Management
- Workload and Project Space Management
- Future, Present and Past based Decision
- Optimize instead of Filter Decisions
- Full Automation instead of Semi-Automation
- Supports Commodity Technologies
Alternate Approaches + Intelligence Engine

Intelligence Engine

- VM-Based
- Workflow/Script-Based
- Peer RM/Fabric-Based
- Mechanism-Based
- Cloud Onramp-Based
- Web Services / Load Balancer-Based

“Intelligence Engine is Additive”
Cloud Evolution
From Semi-Automated Project Spaces to Workload-Driven Real Time Infrastructure

Cloud 1.0
Self-Service Cloud (Project Space)
• Semi-automated
• Self-healing / Static Infrastructure

Cloud 2.0
Workload-Driven Cloud (Real-Time Infrastructure)
• Automated
• Self-healing / Dynamic Infrastructure

Cloud 3.0
Interconnected Ecosystems

A. Applications
B. Application Ecosystems

Public
Private

~ .5x – 2x
~ 3x – 6x

Efficiency Improvement Assumptions

Intelligence Engine
Real-Time Infrastructure

Intelligence Engine

RTI
- Policy-based resource optimization to SLAs
- Policy-based, cross-domain automation
- Policy-based, domain-specific automation

IT Service-Oriented
IT Management Process Maturity

Process Standardization
- Availability Mgmt. Process Domain
- Problem Mgmt. Process Domain
- Change Mgmt. Process Domain
- Provisioning/CM Process Domain

Infrastructure Maturity
Virtualization
Technology Standardization and Consolidation

Gartner: IT Modernization & RTI – December 2008
ID:LSC27_117, page 14, by Donna Scott
### Real-Time Infrastructure and Cloud

**Anatomy of a Private Cloud Service**

**Access Management Tier**
- Self-service interface
- Programmable interface
- Subscriber management
- Identity and access management

**Service Management Tier**
- Service catalog
- Service model
- Service configuration management
- Service-level management
- Service availability and performance management
- Service demand management
- Service financial management

<table>
<thead>
<tr>
<th>Service Governor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource Management Tier</strong></td>
</tr>
<tr>
<td>Resource governor/configuration management</td>
</tr>
<tr>
<td>- Allocation</td>
</tr>
<tr>
<td>- Pooling</td>
</tr>
<tr>
<td>Resource state management</td>
</tr>
<tr>
<td>Resource performance monitoring/usage metering</td>
</tr>
<tr>
<td>Resource security</td>
</tr>
</tbody>
</table>

**Resource Tier (Infrastructure, Platform or Software)**
- Component managers
- Resource pools
- Virtual resources
- Physical resources

---

Gartner: The Architecture of a Private Cloud Service
ID Number: G00171868, 15 March 2010
What is a Service Governor?

A Service Governor can:
• Access Information
• Modify the Workload
• Modify the Environment

A Service Governor:
• Makes and Enforces Intelligent Decisions
• Applies Historical Data & Future Commitments to those Decisions
Unifying and Integrating the System
Workload Driven Architecture
Comprehensive Unification and Intelligence

Core Services
(Workload, Resources, Services, etc.)
- Aggregate State Information
- Unify Services Layer
- Adapt and Optimize Operations
- Autonomic Learning

Orchestrated Services
- Workload Management
- Policy Management
- SLA Management
- Resource Management
- Event and Workflow Management
- Provisioning and Virtualization Management
Intelligence: Context and Policies
Service Governor in the Software Stack

Workload & Application
- Transaction / Workflows
- Batch Workload (Parallel and Serial)
- Other Workload (DB Farm, Web Farm, etc.)
- Home-Grown Application

Service Governor

Resource Manager
- Provisioning Manager
- Hardware Monitor
- Virtualization Manager
- Storage Manager
- Network Manager
- Application / Resource Mgr.
- License Manager

Operating System
- SUSE
- RedHat
- Other
- Scyld
- AIX
- Solaris
- IRIX
- Other Unix
- Mac OS X
- Windows

Resources
- Compute
- License
- Network
- Storage
- Other
Role of a Service Governor

Automation Intelligence

Determine Context
Determine past, current, and predicted context:
- Workload and application demands
- Historical trends
- Future resource reservations
- Resource state (servers, network, storage, licenses, devices, etc.)
- Application and resource health (load, temperature, failures, etc.)

Apply Policies
Apply policies configured by the organization:
- Objectives (business-needs alignment)
- Priorities
- Service-level agreements (Guarantees and Reservations)
- Ownership, security, costing, and billing
- Business continuity

Act Intelligently
Optimize within context and policies:
- Start, move, or modify workload
- Adapt, scale, or repurpose environment (provisioning or VM management)
- Create and enforce workflows
- Activate energy savings (power on/off, thermal balancing)

Unified IT Service
Results:
- Reduced costs
- Faster time to solutions
- Increased service quality
- Rapid response to changing needs

Resources and Workload Needs
Self-Service Portal

Enabling Technologies

Experience Enhancement Capabilities
- Single point of entry for end users
- Packages simplify end-user choices
- Predictability through automation of manual processes
- Fast infrastructure deployment

Failure-Avoidance Capabilities
- Unified world view for scheduling decisions
- Reservation system
- Intelligent placement
IBM Virtual Loaner Program

Key Concept:
Initiates, saves, and restores —
- reservations
- hardware setup
- software images
- configuration
- etc.

Key Benefits:
- Dynamic setup
- Dynamic cleanup
- 3X more usage
- Fraction of cost

Diagram:
- WebSphere Portal
  - VPN
  - SSH
  - Portal View of Available Resources
  - Resource Availability Query
  - Reservation Commitment
- Intelligence Engine
  - Store Reservation Records
  - Query Resources
- Provisioning Manager Tivoli
  - Request Resources
  - Provision Resources
- Database DB2
  - Store Reservation Records
- Custom RM HMC
  - Query Resources
- Middleware Operating System
  - x86
  - AIX
- Persistent Storage
- Software Source

Network Components:
- Storage
- System i
- System p
- 15-way
Workload-Driven Cloud
Enabling Technologies

Experience Enhancement Capabilities
- Single point of entry for end users
- Templates simplify application deployment
- Workload-driven changes to ecosystem
- Dynamic workload balancing

Failure-Avoidance Capabilities
- Unified world view for scheduling decisions
- Hot-spot avoidance
- Reservation system
- Automated failure response with escalation
A Top Global Bank

- One of the world’s largest financial services companies
- Deploying Intelligence Engine and IBM’s xCAT to manage more than 50,000 servers under a single cloud-management engine
- Intelligence Engine provides dynamic intelligent workload scheduling and triggering of automatic resource modification (re-provisioning) to meet workload demands
Advantages of Workload-Driven Cloud

• **Agile:**
  reduce resource delivery time from weeks to minutes

• **Automated:**
  mitigate risk and enforce site-specific rules

• **Adaptive:**
  respond to dynamically changing circumstances according to organizational priorities

• **Cost competitive:**
  3X to 6X efficiencies in workload-driven cloud over traditional IT environments
Invitation

- www.adaptivecomputing.com
- Contact us:
  - Michael Jackson: michael@adaptivecomputing.com,
    +1 (801) 717-3722
- Free technical architecture session