IBM GTS Cirba Case Study


Southern Computer Measurement Group
Fall 2016
October 20th, 2016
Ken Christiance – IBM Distinguished Engineer
Agenda

- IBM Global Services Overview
- Cirba Overview
- Better Sizing Program and Client Stories
GTS Strategy

We are transforming the GTS Services model – shifting from a Systems integration to a Services integration approach to Information Technology in the Enterprise.

From: **Systems** Integration…

- Workforce required for simple, medium and complex tasks
- Emphasis on discrete technical skills

…To: **Services** Integration

**IT as a Service**

- Brokerage
- IBM Services
- 3rd Party Services

**Orchestration**

- Traditional IT
- Private Cloud
- Public Cloud

**Security, Analytics, Automation and DevOps**

**From… skills for systems integration**

- Workforce required for simple, medium and complex tasks
- Emphasis on discrete technical skills

**To… skills for services integration**

- Automation with a highly skilled workforce
- Services-oriented skills, integrated across enterprise

**Upskill for higher value** aligned with the shift to Services Integration (e.g., cross-services and sales consulting, data scientists, industry expertise, 3rd party services expertise, partner alignment and management)
Why is Cirba Relevant for GTS?
By the Numbers…

- 507K Virtual Machines / LPARS Supported
- 40K Host / Hypervisors Supported
- 300 FTEs supporting DS Perf Cap
Transformation to Services Integration....

- Enabling Cloud / Hybrid IT
  - Embedded in Hybrid IT Services
  - Analytics for Placement and Rightsizing
    - Traditional IT. Private/Public Cloud (eg Softlayer)
  - Deployed in GTS Private / Public clouds
- Cloud Brokerage / Governance
  - Optimal placement / best venue selection
  - Gravitant alignment
- IBM and 3rd Party Services Innovation
  - Industry Expertise with Cirba partner alignment
  - Policy-Based Automation
  - Delivered / packaged as a Service
  - Cloud Management Orchestration
    - VRA, ICO, Openstack
GTS Program Overview / Stories

What is Cirba?

- Workload Control Analytics that balance application demand and infrastructure supply to drive efficiency/automation internally and in the cloud

- Partnered with IBM since 2007
  - Standard for Capacity Optimization globally
  - Embedded in DHS (Dynamic Hybrid Services)
Where Does it fit in the Ecosystem?

**CLOUD & SERVICE MANAGEMENT PLATFORMS**

- VMware vRealize
- IBM Cloud Orchestrator
- openStack
- Hewlett Packard Enterprise
- ServiceNow

**API**

**DEMAND MANAGEMENT**
- Route and Reserve.

**SUPPLY OPTIMIZATION**
- Densify and De-risk.

**POLICY AND AUTOMATION**

- KVM
- red hat
- AIX
- Windows Hyper-V
- SoftLayer
- Microsoft Azure
- Amazon Web Services

**HYPERVERVISORS**

- Compute
- Storage
- Network
- Software Licenses

**RESOURCES**
How Does it Work?

- It predictively analyzes workload patterns to optimize workload density.

  Predictively identifying complementary workload patterns **drives higher density**, dramatically reducing infrastructure costs.

  Predictively identifying potential conflict **prevents workload contention** before it occurs, making environments run better.

- VM1 – Busy in the morning

- VM2 – Busy in the evening

- While at the same time reducing workload contention and operational risk
The Importance of Being Predictive

Reactive approaches don’t move VMs until the damage is already done. Resource contention will have already occurred for at least 5-10 minutes, **significantly impacting end users**

(Note: In this example DRS will never act, as it doesn’t look at Disk I/O)

Customer example of the impact of predictive analytics:

Reactive products only look at recent activity, not predictive models:

In the month after Cirba was enabled the number of reactive VM moves during business hours **dropped over 80%**

This is a clear indication of **reduced resource contention**
The Importance of Analyzing Workload Patterns

- CPU Intensive
- Memory Intensive
- Start of Day
- End of Day
The Importance of Analyzing Workload Patterns

Operational Risk
Stranded Capacity

Average Increase in VM Density: 48%
Average Hardware Savings: 33%
Average Software License Savings: 55%

$
How Cirba is Different

- CPU
- Memory
- Disk I/O
- Network I/O
- Overcommit
- Business Constraints
- Technical Compatibility
- Compliance
- Software Optimization
- ...

13
How does it know what “good” is?

- Out of the box policies define “fit for purpose” behavior and hosting
- IBM standard policy settings for Production Critical through Dev / Test
Visualizing Efficiency and Risk

- **Clusters**
- **Hosts**
- **Guests**
- **Risks**
- **Inefficiencies**
Automated Optimization

Freed Up Stranded Capacity

Reduced Operational Risk
Using Analytics to Automatically Route New Applications

New Application or Cloud Request

Windows OS
Requires SQL server
Has customer data
Needs Gold tier storage
Must run on west coast

New York  San Francisco  Toronto  London  Singapore  SOFTLAYER® Bare Metal  Public Cloud
Cirba and Hybrid Cloud Optimization

Off-Prem IaaS

Placement Optimization
Bump-Up
Reclaim

Intelligent Hybrid Routing

Off-Prem Bare Metal

Cross-Provider Routing

Placement Optimization
Bump-Up
Reclaim

License Optimization

Cross-Cluster Workload Routing

On-Prem Virtual & Cloud

Capacity Reservation

Placement Optimization
Bump-Up
Reclaim

License Optimization

On-Prem IaaS

Intelligent Hybrid Routing

Cross-Provider Routing

Cross-Cluster Workload Routing

Capacity Reservation
Next Generation Dashboards

**Risk Analysis**
- Clusters with Insufficient Capacity: 1 of 16
- Hosts at Risk: 17 of 96
- Undersized Guests: 118 of 1,016

**Densification Analysis**
- Densification Opportunity: 2.1x Density Improvement
- Slacking Ratio (VMs/Host): 10.6 vs. 22.7
  - Current: 10.6
  - Optimal: 22.7

**Demand Based Forecast**
- Available Space in Standard VM Units:
  - Today: 316
  - 30 Days: 219
  - 60 Days: 190
  - 90 Days: 192
  - 180 Days: 196

**Hybrid Guest Sizing Accuracy**
- Just Right: 57.5%
- Oversized: 42.5%
  - Guests: 1,179
  - Used Resources: 218 (16.5%)

**Stranded and Available Capacity**
- VM Units: 1,152
- Available Resources: 316 (27.4%)
- Stranded Resources: 300 (26.1%)

**Guest Placement Accuracy**
- Just Right: 39.8%
- Overutilized: 42.7%
- Underutilized: 17.7%
  - Hosts: 96
  - Total VMs: 41
Better Sizing for Workload Demand (BSWD) Summary

**Project Objectives**
1. Improve issue with underutilized server hardware
2. Utilize advanced analytics to identify, recommend & automate capacity changes
3. Provide actionable list to Account Teams for Automated/Manual Execution

**Use Cases / Benefits**

- **Capacity Management – Optimize Supply**
  - Multi-platform (x86 and System p) hardware
    - Setting more appropriate CPU and memory values based on historical data
  - Image placement
    - Optimizing the workload onto fewer physical resources
    - Right sizing to balance vm density and defrag environment based on Policy
  - Software licencing – Cost reduction based on physical location
    - Image placement based on software license models (e.g., Windows, Linux, Oracle)
    - Reduce VMware licensing via hypervisor reclamation

- **Demand Management & Forecasting**
  - Reservation & Routing on Supply; Bursting to Softlayer
  - Potential to replace SRM in the HRM process
Healthcare: Cost Avoidance

Customer

Large healthcare provider in the USA and IBM’s largest Cirba deployment

- Hypervisors: VMware & PowerVM
- Servers: 31,000 VMs/LPARS & Servers
- Cirba Customer Since: 2015

Challenge

Requires a solution that provides;

- Ability to optimize server environment to reduce Capital and SW expense;
- Increase system density and reduce / eliminate unused capacity within environment via reclamation;
- Address systems Performance Risk due to Capacity Constraints
- Improve planning and governance model for new Hybrid Workloads

Solution: Densification, De-Risk, Automation & Hybrid IT

- Cirba’s workload optimization analytics was used to identify 20% infrastructure recovery with VMWare
- 4700+ automated rebalance actions applied
- Establish reclamation governance process for right sizing virtual workloads and remove operational risk
- Expansion of Cirba capabilities for ICO routing and reservation of Cloud workloads

Measurable Success

- Reduction in costs to tenants relative to legacy environment > 20% (VM) > 20% (AIX) *
- Automated rebalance activities that reduce operational risk for hypervisors > 4700 (VM)
- Reclamation benefits identified to date (conservative policies) > $2M *

* Estimated
Customer
Government provider of IaaS cloud services that address data sovereignty requirements in a “pay as you go” model

- Hypervisors: VMware
- Servers: 2,000 VMs & servers

Challenge
Cloud needed a solution that:
- optimizes cluster capacity to predict and defer infrastructure spend;
- provides tenants visibility to workload performance and VM right-size recommendations;
- and streamlines capacity management overhead to efficiently oversee the environment, which is expected to grow dramatically

Solution: Densification
Cirba’s analytics were used to densify the virtual environment and identify true infrastructure requirements. Reporting on tenant VMs showed customers which VMs were running “at risk” (especially for memory usage) and needed to be upsized.

Measurable Success
- > 40% Reduction in costs to tenants relative to legacy environment
- > 10% Increase in revenue due to upsize recommendations from Cirba
- > 50% Effort savings for capacity planners to manage the cloud environment
Customer
IBM Managed Infrastructure for energy company engaged in global production of crude oil and natural gas.
- Hypervisors: VMware
- Servers: 1750 servers

Challenge
The IBM team managing the client’s infrastructure wanted to:
• Improve visibility into operational risks
• Accurately forecast and model capacity requirements
• Leverage existing infrastructure more comprehensively

Solution
✓ Cirba’s analytics reduced operational risks through VM rebalancing to eliminate over-stacking and load imbalances. In addition, CPU and memory allocation changes reduced the risk of VMs being starved for resources.
✓ The team also identified 54 nodes they wanted to migrate to 3 specific clusters, using Cirba to identify the appropriate placements for workloads. The team saved a minimum of 2 to 3 days that would have previously been spent gathering CPU, memory and storage details to factor into the plans.
➢ “[Without Cirba] We would have absolutely no way of knowing whether we could successfully move those workloads to other clusters without introducing risk into clusters that were risk free before we added the additional workload.”

“For me, it's all about what Cirba gives us now…The ability to make informed decisions about workload placement and CPU/RAM allocation and de-allocation”
Chemical: Reducing risk

Customer
Large chemical company.
Servers: 500 AIX on PowerVM, 2500 VMware

Challenge
The IBM team turned to Cirba for cost savings through improved density and right-sizing as well as enabling more accurate demand management and forecasting.

Solution
✓ Cirba automates infrastructure optimization for the IBM team to ensure ongoing efficiency. The analytics dashboard also enables central demand management and forecasting to streamline and improve resource planning.
✓ While Cirba has not been released into production, early analysis has identified a number of opportunities to reduce risk and increase efficiency while preparing for cloud within the mix.

“Using out of the box settings, the DPE and I were presented with a realistic action plan for right sizing clusters, host, and virtual machines in our environment.” Chief Architect for the Account, IBM Certified Executive Architect

- 44% Of environment has densification opportunity
- 17% VMs over-sized, opportunity for right sizing
- 11% Hosts overloaded
- 14% VMs starved for resources
thank you!