The Dynamic DataCenter

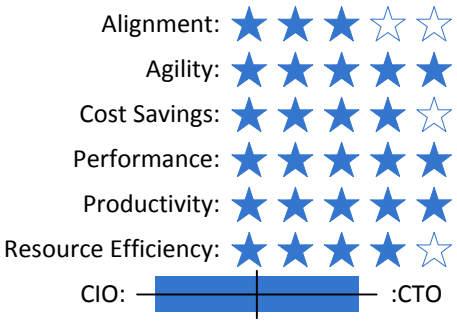
Fit-for-Purpose Design with Real Time Infrastructure

January 2009

Enterprise IT is in the midst of an evolution in the way the applications, information, content, compute, storage and network is deployed and consumed. As things stand today, costs are out of control and it is difficult to relate growth to cost in the datacenter; the complexity is becoming unmanageable and there is poor alignment between the business and its supporting technology.

The move is toward an efficient, reliable and tailored operating platform. The Dynamic Datacenter distinguishes the service delivery of IT. It respects the relationship between lines of business and their collaborating applications. The migration toward this dynamic and real-time service oriented infrastructure is based on business demand and allocated infrastructure supply to meet that demand as needed / when needed, just like a utility.

Executive Summary



The business is demanding that IT become an enabler of competitive advantage. This competitive advantage takes many forms including responding to business events in real time, reducing product time to market, enabling higher transaction volume/complexity, shrinking execution cost, etc. IT has fallen short of achieving these goals with the general approach of optimizing a few critical applications. For IT investment to realize its full potential, a more holistic approach is necessary: a fundamental shift from "Deploying Application Infrastructure" to "Deploying Applications onto Allocated Infrastructure", which requires that an infrastructure be optimized for the types of work an organization executes (Fit-for-Purpose) and is dynamically/rapidly allocated on demand (Real Time Infrastructure).

Business Drivers / Strategy

- **Simplify the Delivery of IT:** The processes and complexity of deploying infrastructure and applications often takes weeks or months and directly affects business profitability.
- **Agility:** Beyond reducing the time to deliver supply more quickly, the business needs to dynamically scale in real time to have resources where / when needed.
- **Performance:** Resources are typically picked from what is pre-selected versus what is optimal, which is inefficient capital usage.
- **Environment Efficiency:** Even with the recent attention to consolidation programs, large portions of IT infrastructure are poorly utilized, which directly affects the profitability of the business.

Key KPI Improvements

- **Time to Market:** up to 50+% reduction
- **Total Cost of Ownership:** up to 60+% less in Capital and Operating Expense
- **Performance:** up to 50+% improvement
- **Datacenter Productivity:** up to 70% reduction in packaging, deploying and changing applications and infrastructure

Execution Strategy

- **Virtual and Dynamic:** Allocate low overhead virtual resources from physical ones dynamically and in real time.
- **Fit-for-Purpose:** Creation of Processing Execution Destinations (PEDs), where the characteristics of the workload are matched with the ability of the resources.
- **Robustness and Performance:** Since the resources are virtual and dynamically allocated, replacing failed resources or scaling up environments as needed becomes significantly easier.

Key Technologies

- Partner: Cisco
  - ACE XML Gateway
  - Enterprise Policy Manager
- Partner: DataSynapse
  - Dynamic Application Service Management (DASM)
- Partner: Scalent
  - Virtual Operating Environment (VOE)
- Partner: Oracle
  - Coherence
- Partner: VMWare
  - Infrastructure

The focus of the paper is describing a Fit-for-Purpose Real Time Infrastructure to address today's infrastructure challenges. The approach is radically different from traditional approaches to managing infrastructure. A review of tradition approaches and their challenges is followed by a detailed description of the proposed solution and how to operate that solution.

- **Problems with Traditional Infrastructure Approaches**—Challenges with current implementations and the related consequences.
- **Proposed Solution: A Fit-for-Purpose Real Time Infrastructure (RTI)**—An Innovative strategy to implement a dynamic infrastructure that resolves the challenges of traditional approaches.
- **Realizing a Fit-for-Purpose RTI Design**—How to turn the vision into a reality.
- **Sustaining and Governing a Fit-for-Purpose RTI Design**—How to create lasting organizational value.

Building a RTI requires a set of technologies and practices that are structured through a methodical implementation approach. We describe the key business and technical aspects of building a Fit-for-Purpose RTI that enable IT to support business in a more aligned manner, while being more agile and efficient.

Problems with Traditional Infrastructure Approaches

Both providers and consumers of IT infrastructure are frustrated by the complex and time consuming process of deploying infrastructure. Once resources are placed, they are by nature very rigid and changes are difficult to implement, and two fundamental problems exist.

1. The traditional approach to design, procurement, deployment and management activities is not holistic; instead the tasks are separated into silos of control.

    - Application tiers (e.g. middleware, database) are not usually placed in close proximity, causing added latency.
    - Application growth may force new dedicated servers to be dispersed across the datacenter, further slowing response times.
    - Highly collaborative applications are rarely placed in close enough proximity, which increases latency, affecting response time and overall performance.
    - Movement of hardware resources around datacenters is currently unacceptable due to a lack of instrumentation to gauge and report on impacts, not to mention the potential for interruptions of service.

2. Design and procurement are based on individual application need, and are typically marshaled through an organization by the application team.

- The team usually designs the hardware to process 2–5 times current anticipated peak load, leaving the resources significantly underutilized at most times. But, because of the inherently fixed configuration, no other application or service can share that unused capacity.
- Metrics and monitoring tools are rarely robust enough to report, anticipate or react to change; therefore, the consequences of spikes in demand are rarely correlated across the entire application.
- Deployment and management fall under the control of Operations and/or Engineering where configurations and physical placement are arranged by hardware asset type (e.g. database servers have a section of the datacenter, application servers are placed in a different section).
- The introduction of solutions utilizing virtualization and varied dynamic technologies are treated in the context of a single (or few) application(s); therefore, only a portion their value is captured.

While each compromise in and of itself doesn't appear significant, the result is a datacenter full of applications and their oversupplied infrastructure that are randomly disbursed creating a maze of physical and logical interdependencies, destroying value due to losses of efficiency and agility. For example:

- Server growth is uncontained, but utilization on a 24-hour scale is typically less than 20%.
- Power, real estate and personnel costs continue to grow unabated.
- The growth in servers and other infrastructure does not guarantee meeting SLAs for critical applications.
- Rapid growth of a datacenter increases complexity and cost while often degrading performance and availability, two key measures of user experience.
- All measures of Quality of Experience (user experience, cost and efficiency) are subject to failure on a regular basis—balancing them becomes a full time job in futility.

The end result is that businesses with enterprise datacenters are facing considerable pressure. A new approach to managing datacenter infrastructure is needed to alleviate these problems.

Proposed Solution: A Fit-for-Purpose Real Time Infrastructure (RTI)

A Fit-for-Purpose RTI represents a shift in thinking about how total IT supply is designed, purchased deployed, managed and utilized, which facilitates the elimination of silos of operation and the management of the environment holistically; the primary challenges of the traditional approach.

This solution is comprised of two fundamental concepts: Fit-for-Purpose and RTI. While each concept does not require the other to operate, together their synergy creates significant business value: a dynamic and real-time service oriented infrastructure based on business demand that allocates optimized resources from infrastructure supply to that demand as needed / when needed. Let's explore these concepts further.

What is Fit-for-Purpose

For a moment, instead of resources in the datacenter, let's consider personal transportation; specifically the following categories of vehicles: sports cars, SUVs, pickup trucks, and U-Haul trucks. With little thought you'd be able to pick an appropriate vehicle for most any situation. A weekend getaway trip?—the sports car. Taking the kids to soccer practice?—the SUV. Helping a friend move out of an apartment? Though the pickup may work, the U-Haul truck is clearly superior. The point is picking the right vehicle category for the trip can make life easy or difficult. Fit-for-Purpose is about avoiding ending up with landscaping stones and mulch in the trunk of your sports car.

Resources in a datacenter are often deployed without consideration to the work characteristic at hand. The demands of work can vary significantly and can be broadly categorized into the following major categories: Low Latency, Complex Transactions, Numerical Processing, and Information Integration. Rather than force fitting work onto generic hardware, it's important to consider the needs of each work category and how to optimize supply to them. For example, the creation of a Java Optimized execution environment with a Java Appliance Accelerator is a significantly more effective solution than each application deploying another application server to run Java.

What is Real Time Infrastructure (RTI)

The purpose of RTI is to implement fluidity of resources to meet demand in real time while enhancing Quality of Experience[i] and enabling transparency of processing so that latency can be understood for every component of an application and to facilitate faster deployment time providing a better competitive edge. This requires autonomic management capabilities that predict problems and enable self healing in case of failure of a system, a global dashboard view that relates business transactions to its effect on infrastructure supply, and guaranteed execution of designated transactions based upon business policy. For a sampling of the different types of technologies and products employed in an RTI solution, see Figure 1 - RTI Core Elements and Technologies.

Operating Characteristics of a Fit-for-Purpose RTI

Putting these two concepts together dramatically improves their individual value. Real Time Infrastructure is designed to make IT as agile and responsive as the business needs. While RTI alone does yield agility, there is a limit to its ability to generate efficiency. The addition of Fit-for-Purpose acknowledges that each resource and element of work has different capabilities and characteristics. A Fit-for-Purpose Infrastructure matches resource capabilities with work characteristics for more efficient and better execution. Together, the concepts of RTI and Fit-for-Purpose work in concert to create multiple pools of resources that can be dynamically brought to bear for the various types of work the organization is challenged to execute, which facilitates holistic operation and management of the environment.

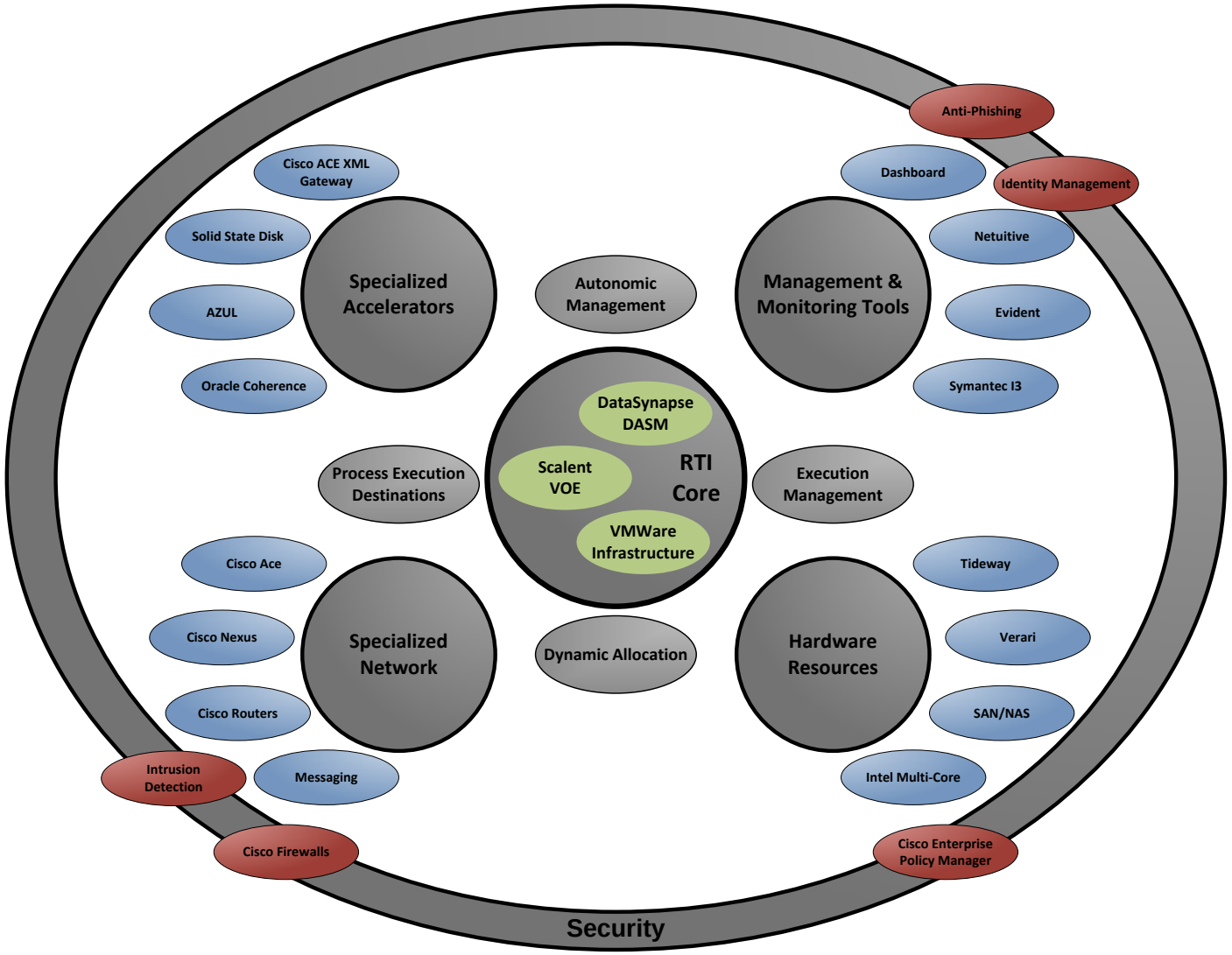


Figure 1 - RTI Core Elements and Technologies

The key operating characteristics of Fit-for-Purpose RTI are:

- **Dynamic Allocation** of IT supply to real-time demand based on service contract requirements, including guaranteed execution response, throughput volume, work characteristics and time constraints or cost / margin rules.
- **Real-time Transaction Work Management** that synchronizes work characteristics (demand) with resource capabilities (supply) to meet execution management service contract requirements.
- **Self Healing Attributes,** such that failures are seamlessly repaired, without human intervention.
- **Efficient Execution of Work** on infrastructure that is tailored to the type of processing needed.

Realizing a Fit-for-Purpose RTI Design (Design Principles for RTI)

Realizing a Fit-for-Purpose RTI Design requires a significant understanding of the needs of the environment. Here are the four key areas that need to be addressed to realize the design, which will be elucidated in subsequent sections of the document:

- **Drivers:** The business and IT drivers that define the goals of the design.
- **Artifacts:** The quantification of business consumption demand and platform infrastructure supply that describe the usage of the environment.
- **Core Competences:** To take advantage of the understanding formed from the artifacts, there are a number of core competences that need to be developed.
- **Technologies:** With the above understanding and competences, technologies must be deployed to realize the vision. The Artifacts are used influence the selection from the various available technologies to select those that best meet the Fit-for-Purpose design principles.

These breakdowns are followed up by a **Step-by-Step Guide** for executing the design.

Drivers

Additional benefits beyond solving the traditional challenges are enabled by this type of design. To realize its full value we need to consider the entire spectrum of business and IT drivers. Note that there are two sets of goals: those of the business consuming the infrastructure (demand), and those of the IT Infrastructure team (supply). Both must be considered in developing the design.

Business Goals

- **Reduced Time to Market:** IT becomes agile enough to add new services and products, reducing lead time from months or weeks to days.
- **Business Capacity:** IT can increase its processing capability as the business grows, handling unexpected spikes in transactional demand in real time.
- **Budget Predictability:** IT can show that its costs are predictable and proportional to business growth.

- **Competitive Advantage:** When fully realized, RTI represents a significant step toward IT becoming a strategic asset for the business instead of just another cost center. IT becomes a strategic enabler for the business.

IT Goals

- **Supply Efficiency:** RTI mandates a fluid set of infrastructure services that make resources available to meet oncoming demand at the level that is needed, when it is needed.
- **Utilization Efficiency:** RTI assumes demand fluctuates for different applications at different times.
- **Demand Driven Utilization:** RTI utilizes a Fit-for-Purpose analysis approach that promotes a top-down demand-driven design approach that can be configured bottom-up for specific application needs.
- **Reuse:** RTI respects and incorporates physical limitations into the Fit-for-Purpose analysis, where the operational characteristics of performance, security and availability are compared across the constituent patterns of applications to determine similarity.
- **Performance through Localization:** RTI principles include consideration for proximity when assembling an execution environment for collaborating applications.
- **Performance through Optimization:** RTI designates four fundamental operating environments (infrastructure patterns) that facilitate specific operational behaviors. Any application once analyzed will be able to operate in one of those four environments.
- **Quality of Service:** RTI provides a tactical capability to utilize resources based on business polices that balance: user experience, efficiency and cost.
- **Improved Processes:** RTI tactics affect how systems and infrastructure are designed, managed, serviced and utilized.
- **Organizational Roles:** RTI tactics positively affect how IT organizations are structured in terms of roles and responsibilities, providing more agile response to new needs and problem resolution and determination.

Artifacts

Once the drivers are well understood, a number of artifacts are developed to characterize how the environment should be utilized. The purpose of these artifacts is to quantify the environment in terms of Business Consumption Demand and Platform Infrastructure Supply, which facilitates a healthy and holistic grasp of the complexity of the problem being solved from the two primary perspectives of the business consumers and the infrastructure supply team. The following descriptions are of artifacts that are generally created in the process of developing, deploying and using a Fit-for-Purpose RTI environment.

Business Consumption Demand

- **Business Value Chain:** The creation of a value chain depicting how business execution correlates to IT infrastructure implementation of underlying capabilities.
- **Forensic Analysis:** A process whereby the existing business platform and supporting infrastructure are identified and holistically documented. These details are usually left to local system managers, so organizations frequently have a lackluster overall view of existing systems, their purpose, and how they interact. This information is critical in planning a future state business platform.
- **Current State Alignment Map:** A number of views used to describe how the Business Value Chain maps to products and promotes the desired Qualities of Experience (QoE), including latency, throughput, resiliency, and cost. Once available, infrastructure gaps and duplicate systems in the value chain become apparent. Their elimination makes the organization more efficient and effective with its IT budget.
- **Product and Portfolio Management:** Coalescing business and technical priorities in continuous capability adoption to ensure sustainment and differentiation of your business through IT.
- **Economic Model:** Define the business and IT linkage of demand and supply. Orient analysis and model output around the interactive dynamics of consumption of IT resources by the business and the fulfillment behavior of processing by IT. This needs to be correlated with the value chain function and the corresponding products or services that are differentiated by business type (liquidity, risk transference, advice), business importance (margin, labor, flow) and cost to transact.
- **Dynamic Economic Model:** A defined business and IT linkage of demand and supply. The model must be created around the interactive dynamics of IT resources by the business and the fulfillment behavior of processing by IT.

Platform Infrastructure Supply

- **Demand Mapping:** In natural language terms (no geek speak!), define and capture the "day in the life of the business", what it expects and where there are problems. Be sure to understand sensitivities to cost, bottlenecks and timing constraints as well—specifically, quality attributes and operational requirements of the business in terms of calendar events (e.g. different usage at end of month), demographics (i.e. user types, business lines, and geographic spread), and competitive benchmarking across the straight-through-processing (STP) value chain.
- **Future State Design Map:** Infrastructure gaps, missing characteristics, and duplicate systems are identified from the Current State Alignment Map. The Future State Design Map depicts the resulting business platform once those issues are resolved. This results in additional efficiency and more effective monetary allocation to IT.
- **Fit-for-Purpose Operational Footprints:** The categorization of operational qualities based on demand facilitates the realignment of processes, thereby optimizing infrastructure resources to required characteristics. Tailored and coordinated operational footprints are organized in tightly coupled physical environments called Processing Execution Destinations (PEDs).

- **Abstraction as a Fundamental Design Principle:** The design of each infrastructure component and layer needs sufficient abstraction so that operational details are hidden from all collaborating components and layers. This maximizes the opportunity to create a virtualized infrastructure, which facilitates rapid deployment and simplicity of operation.
- **Service and Change Management:** Dynamic models of operation require processes and procedures of service, delivery and change management to accommodate the "on-the-fly" and "as needed, when needed" operations of IT.

Core Competences

Having drivers and artifacts are not enough to develop the full understanding required for designing and efficiently utilizing the environment. A number of core competencies are needed from those involved, including a solid sense for and ability to perform:

- **Business and IT Alignment**: to understand high impact areas, biggest pain points and future state vision.
- **Dependency Mapping:** to gain insight in to the inter-relationship of applications, content and services.
- **Rigorous Measurements Discipline:** to discover demand and supply characteristics.
- **Fit-for-Purpose Design:** to create optimal platforms based on demand.
- **RTI Implementations**: to pilot, verify, validate and educate.
- **RTI Operations**: to ensure consistent resilience and efficiency.
- **RTI Sustainable Model**: to create and maintain a strategic asset for the business.

Technologies

In the context of Business and IT Drivers, with the understanding of the business consumption demand and platform infrastructure supply artifacts, and by the abilities of the core competencies a Fit-for-Purpose RTI technology design can be created/implemented that meets the business needs and yet be manageable by IT. The following list contains a number of the technologies commonly used and their primary value, followed by an example of how they have been brought together in a Financial Services environment.

- **Grid Brokers:** Dynamically allocate workloads to resources with appropriate processing and memory characteristics.
- **Service Brokers:** Dynamically allocate work in real time to available compute resources based upon business policy. (See Figure 2)
- **Provisioning Tools:** Prepare application "personas" that create virtualized infrastructure resources. (See Figure 2)

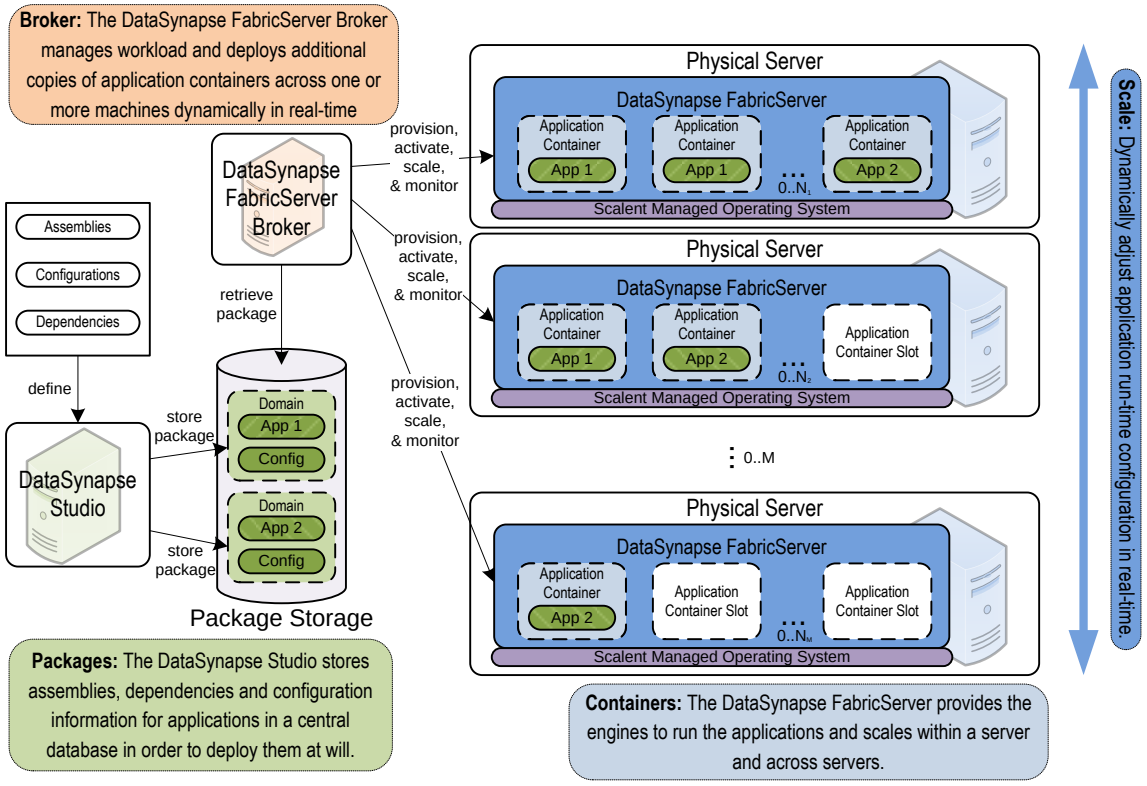


Figure 2 - Scalent VOE™ and DataSynapse FabricServer™: Manage current and new resource infrastructure dynamically. VOE provisions bare metal and FabricServer enables the ability to define, package, provision, activate, scale and monitor applications dynamically adjusting where and the number instances of an application based on demand—seamlessly and in real time.

- **Network Attached Appliances:** Offload intensive processing from servers, enhancing throughput while significantly lowering power consumption and datacenter footprint utilization.

- **Network Accelerator Devices:** Advance protocol switching, streamline authorization and format translation, enabling faster less complex integration.
- **Solid State Disks:** Provide large data set access at memory access speed.
- **Transactional Caching and Data Federation Facilities:** Enhance performance, reduce data movement and facilitate transparent multi-source data access.
- **Instrumentation Technologies:** Measure latency across all hops a transaction takes correlated measurement of an application across all tiers, comprehensive utilization dashboards and predictive analysis tools.
- **Optimized Processing Containers:** Reduce power consumption and floor space by utilizing unique vertically cooled processor racks. (See Figure 3)
- **Multi-core CPUs:** Accelerate multi-threaded applications and consolidate processing of single threaded applications. (See Figure 3)

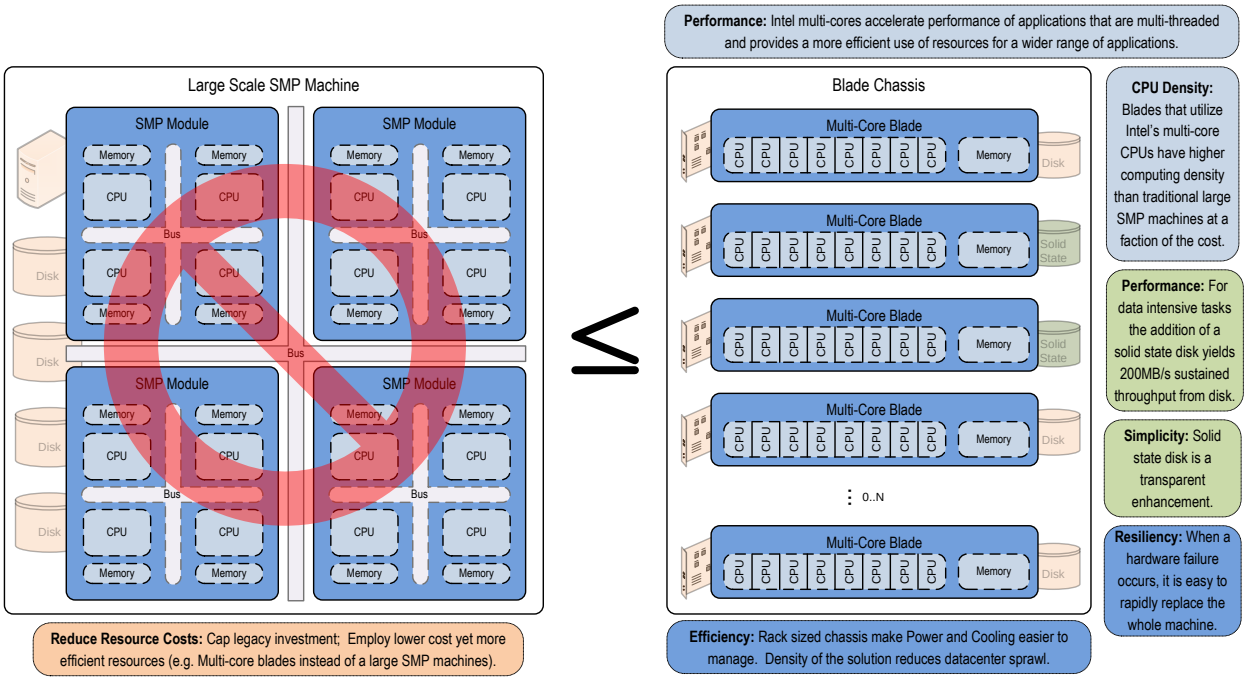


Figure 3 - Multi-Core CPUs with Solid State Disk: Cap legacy investment; Employ lower cost yet more efficient resources. Blades that utilize Intel's multi-core CPUs have higher compute density at a fraction of the cost. Additionally a blade solution has more flexibility on how it is deployed than traditional large SMP systems.

- **Specialized CPUs:** Application logic burned into a chip (field programmable array), that runs many times faster than conventional software.
- **Dependency Mapping Facility:** Presents facts about how applications and services are related in production.

An example of pulling all these technologies together for a Financial Services Risk Functions is presented in Figure 4.

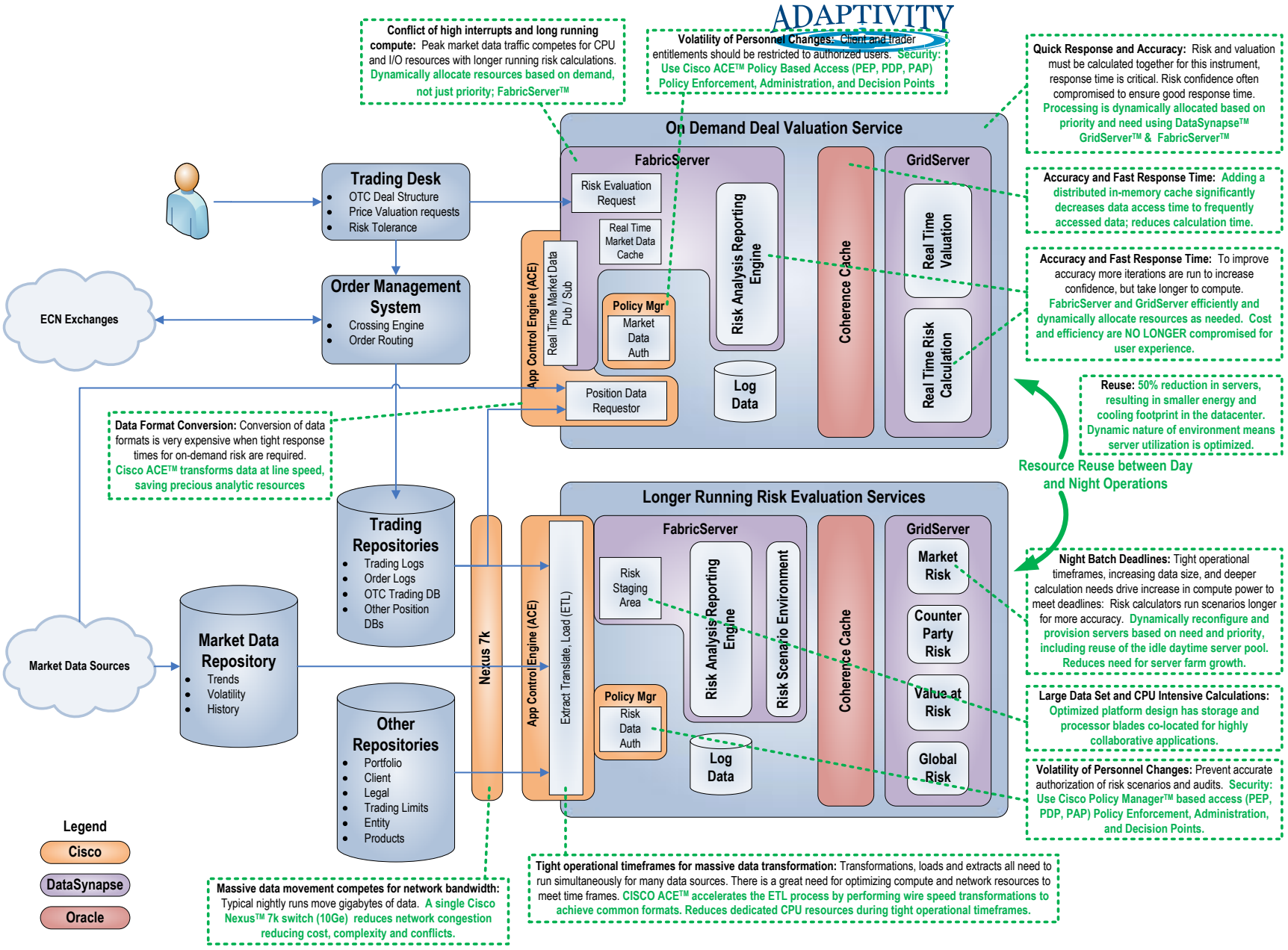


Figure 4 - Financial Services Risk Functions optimized to run on a Fit-for-Purpose RTI Platform

Step-by-Step Guide:

Achieving RTI requires a journey that takes the organization through stages that mature certain processes and people skills while introducing new enabling technology that makes RTI sustainable. A particular implementation of RTI doesn't end when a new set of technologies is introduced because the organization's business goals change, the marketplace shifts, and when RTI is established correctly the business can engage capabilities and envision strategies that were not possible in more traditional operating environments. The major stages of this journey involve the ADIOS framework.

- **Alignment** of business goals and realities with IT tactics that can properly support those goals. Activities include understanding the current demand profiles of critical applications for key lines of business, the current IT and business pain points, along with a vision for the optimal operating environment that will support business goals.

- **Design** of a dynamic demand-driven platform that can optimally operate classes of applications by business function and operational characteristics. This incorporates an in-depth understanding of the types of business demand placed on applications and the constraints of current implementations.

- **Implementation** of these tailored platforms where the integration of sets of enabling technology is emphasized. These new technologies require process changes and skill augmentation, and are migrated incrementally, along with better monitoring solutions that provide a more accurate depiction of how the new and old environments are operating to meet demand.

- **Operation** of the new platforms which require changes in thinking about the on-boarding process for new applications, new approaches for day-to-day operations, the use of the integration lab and the types of roles, responsibilities and shared resource billing needed.

- **Sustainment** of the platform which is the key to long-term success. This mandates that the measurements of the environment are actively used in future planning, that investments in IT are viewed from a Portfolio Management perspective where ROI is gauged, and investments are subject to governance. This requires a shift in thinking about IT. IT transcends its cost center status and becomes a strategic enabler of the business.

The following guide to the development of a Fit-for-Purpose RTI environment is based upon a case study performed at Wachovia CIB, where the above mentioned phases were employed and structured into actionable steps, with specific objectives and outcomes.

NOTE: Case study specific examples for each step are noted below the step in the following format: Info.

Align

1. Define the Business Value Chain to understand the key lines of business and the associated functions that drive revenue and IT consumption.

   At Wachovia's Corporate Investment Bank (CIB) the entire Fit-for-Purpose RTI strategy was established by first understanding the breadth of the application inventory across all business lines (Capital Markets, Fixed Income, Treasury Services, International, M&A, Risk Processing, Wholesale Operations, etc.) In numerous workshops, one line of business (LOB) at a time was focused on identifying the business value chain (e.g. sales, trading, operations for Capital Markets, etc). These value chain components are the actual functions that the business performs to generate revenue and defray risk. Each LOB described the suites of business products they support (e.g. equities, exchange traded derivatives, OTC derivatives for Capital Markets). These critical business foundations provided context for the placement of all applications for a LOB. Every application could be understood in terms of its relevance to revenue, risk, reporting and compliance, establishing relative importance for a LOB, and also providing the foundation for investigating redundant functional overlaps that should be consolidated.[1],[2]

2. Create a current state alignment map indicating business defined operating requirements for key applications.

   Once the context of aligned applications was established, the business demand that these applications must meet was established in that context (e.g. throughput volume, where low latency matters, tight operational time windows, perceived trends in peak volume). For CIB, this established clear priorities and clearly identified pain points that would be the first focus of various measurement and analysis exercises, including fixed income trading and OTC derivatives risk analysis.[3],[4]

---

[1] A New Foundation: SOA Implementation for Business Transformation
A successful services-oriented architecture, or SOA, implementation starts with making the right decisions early on in the process.
By Nancy Feig, Wall Street & Technology, April 26, 2007

[2] Case Study: SOA Delivers for Wachovia – With Interest
Align Online, November 14, 2007

[3] Tech Earns Interest at Wachovia, eWeek.com

[4] Wachovia SOA Married to The Process
By Michael Hickins, InternetNews.com, May 16, 2006

3. Perform inventory and dependency mapping on key applications discovered through the business value chain exercise. In this step highly collaborative application groups are defined.

   To validate management's perception of application interrelationships and the consequences of those interactions on IT supply, a mapping and discovery tool was utilized to determine the actual physical relationships of the application sets. This helped determine whether the connections between applications and services had been compromised over time due to datacenter sprawl. This enabled CIB to plan how any new initiatives would be organized in the datacenter but also identified key bottlenecks in fixed income trading processing, which were costing the company millions of dollars per day. It also illustrated how efficiencies could be best obtained through movement of machinery, as opposed to new investments, which would not have completely solved the bottlenecks.[5]

Design

4. Run consumption monitoring for definitive metrics on current environment operations.

   In order to determine how demand is actually met during peak time, various approaches to consumption monitoring had to be invoked. This included implementing a real-time transaction latency tagging tool, so that a transaction's latency could be tagged on each leg of its journey. That is how CIB found out the cause of a 10 second delay in fixed income trading. In addition, other tools needed to be introduced that would gauge the affects of demand on an application across all tiers that the application runs on.[6],[7],[8],[9],[10]

---

[5] Business Process Transformation with Virtualization Oriented Architecture
ALIGN 07 – DataSynapse User Conference, Sheppard Narkier

[6] A New Foundation: SOA Implementation for Business Transformation
A successful services-oriented architecture, or SOA, implementation starts with making the right decisions early on in the process.
By Nancy Feig, Wall Street & Technology, April 26, 2007

[7] Red Hat Application Profile for Datasynapse FabricServer

[8] The serverSide.net, Tangosol extends data grid support to Microsoft .NET, Posted by: Michele DiSerio on March 20, 2007 DIGG

[9] Waters, NETWORKS, For the Greater Grid
Wachovia's grid network goes cross-asset and cross-business, By Bob Giffords

[10] Aligning business and IT through instrumentation
InfoWorld Blog, By Tony Bishop, March 22, 2008

5. Decompose the key application workload characteristics against alignment map and monitoring results; application behaviors are targeted for one of the RTI platforms.

   CIB used results from its initial monitoring and mapping along with focused workshops to validate why applications operate as they do. This enabled the RTI team to focus on operational bottlenecks with application teams and operations staff. It pointed out what infrastructure alignments needed to be changed, what made sense to invest in, and the best approaches for bottleneck resolution.[11],[12]

6. Establish forensic analysis to determine gaps in current infrastructure.

   While the workshops were useful, the RTI team needed to distribute knowledge so that operations could take responsibility for running forensics, especially when trouble arose. It was critical for CIB to have a forensic analysis process as part of its day-to-day operations given that much of the infrastructure was in flux. The analysis began to show gaps in under utilization, clearly identifying that while utilization over a 24-hour period was less than 20%, there were obvious spikes across wide areas of the infrastructure where utilization exceeded 90% for an hour or more during the trading day.[13]

[11] A New Foundation: SOA Implementation for Business Transformation
A successful services-oriented architecture, or SOA, implementation starts with making the right decisions early on in the process.
By Nancy Feig, Wall Street & Technology, April 26, 2007

[12] NetworkWorld, Virtualization hits the big time
Deploying virtualization everywhere increases benefits but creates management headaches
By Joanne Cummings, Network World, 11/11/06

[13] Split-Second Securities Trading at Wachovia
Wachovia's service-oriented architecture triples processing capacity at one-third the cost. May 21, 2007 (Computerworld)

7. Understand the consequences of meeting demand by decomposing the supply gaps.

   Some of the gaps were not easy to isolate, such as occasions where large file movements due to testing occurred and interfered with the trading network. Many of the incidents were consequences of point solutions that add network hops between collaborating systems.[14],[15]

8. Create the economic model based on business demand, available supply and importance to the business; this guides the migration process to the RTI platform.

   At CIB it was important to present a model that considered driving business forces. In Capital Markets, latency and throughput were critical drivers and therefore infrastructure investments were needed to accelerate them. Normal disaster recovery (DR) standards were not applicable, so special arrangements for failover had to be accounted for in the cost structure. This was compared with the potential for revenue loss.[16]

9. Instrument transactions in real time so that the changes that will occur as a result of the RTI platform can be properly gauged.

   CIB established a baseline for key transactions in terms of latency, response times and peak utilization of resources so that when the investments started to roll out, they could be measured to gauge the degree of effectiveness especially in key areas such as risk and trading. Early baselines established the case for a redesign of a key risk platform. The original plan was to spend $6 million, taking one year to implement. Early baseline statistics indicated that this platform could benefit from RTI and save $5 million and six months of effort.[17]

---

[14] BANK, Systems and Technology, Wachovia's Global Trade Services Group, under the command of Hong Kong-based Steve Nichols, has deployed workflow and imaging technologies in order to master the complexities of the trade finance business.

[15] Split-Second Securities Trading at Wachovia
Wachovia's service-oriented architecture triples processing capacity at one-third the cost. May 21, 2007 (Computerworld)

[16] A New Foundation: SOA Implementation for Business Transformation
A successful services-oriented architecture, or SOA, implementation starts with making the right decisions early on in the process.
By Nancy Feig Wall Street & Technology April 26, 2007

[17] Aligning business and IT through instrumentation
InfoWorld Blog, By Tony Bishop, March 22, 2008

10. Define the future state alignment map where high value applications and infrastructure gap remediation are all placed in priority order to guide the migration to RTI.

    This is the beginning of an investment strategy approach, where high-value and high-impact changes are prioritized. CIB was able to identify planned projects across Capital Markets, Risk and Fixed Income, that touched on shared services (e.g. risk, position), shared frontends (multi-asset trading frontend), server frameworks, and a host of technology initiatives (e.g. network attached processors, network appliances, solid state disks, transactional caching). All of these were tied to specific remediations with pain points and value propositions.[18],[19],[20]

11. Fit-for-Purpose design is established where applications, characterized by pattern type, are defined as candidates for specific RTI operating platforms through a combination of forensic efforts and application profiling. This effort establishes what applications would benefit from service brokers, network attached processors, network accelerators, and solid state disks.

    The design phase included a rigorous design tradeoff process which manifested in the creation of four sets of integrated ensembles and a suite of defined service layers that would use those ensembles. These ensembles covered the breadth of application families displaying the following key operating characteristics (low latency, complex transaction, information integration, and numerical processing). These four ensembles were designated to service a separately designed client access framework which would interact with the ensembles through a separate mediation framework. The ensembles would run on a set of tailored operating platforms and would access a persistence utility.[21]

---

[18] A New Foundation: SOA Implementation for Business Transformation
A successful services-oriented architecture, or SOA, implementation starts with making the right decisions early on in the process.
By Nancy Feig, Wall Street & Technology, April 26, 2007

[19] Wachovia Delivers High-Performance Computing Through a Services-Oriented 'Utility'
In a pioneering approach, Wachovia is becoming more and more discriminating about how and to whom it provides high-performance computing.
By Penny Crosman, Wall Street & Technology, March 19, 2007

[20] Wall Street-Style Power, With high-performance computing strategic and costly, companies push efficiency
By Penny Crosman, Wall Street & Technology, InformationWeek,
Apr 14, 2007 12:00 AM (From the April 16, 2007 issue)

[21] Hands-On SOA Governance Practices and Processes
Information World SOA Executive Forum (5/16/2006), Sheppard Narkier

12. Establish the RTI lab where the initial technologies are integrated and candidate application patterns are tested.

    The lab had two purposes, run rigorous integrations of current and new technologies, making it easier for the framework teams to deliver functionality, and provide a sandbox for development teams, many of whom were skeptical of the functionality or scared of the learning curve. This lab was also able to demonstrate how dynamic allocation of resources worked in a practical manner that spoke to developers.[22]

Implement

13. Introduce applications that run on service brokers to optimize resources based on real-time demand and business policy.

    One of the quickest wins for CIB was to take existing applications that ran in J2EE environments and more effectively allocate the containers to supply based on real-time demand. No code change was required when employing a service broker to better utilize existing resources and handle spikes in incoming online requests for browser-based applications in target areas across CIB's LOBs. The service broker was delivered as part of a framework offering that included caching and object relational mapping.[23],[24]

14. Introduce network attached processors that accelerate java processing: migrate candidate applications.

    Even with service brokers, java applications have inherent limitations that prevent memory sizes from growing above 2GB. Garbage collection pauses are unpredictable for java applications and affect highly interactive applications the most, because they have greater memory volatility. Network attached processors (NAPs) require almost no code changes, but provide an operating environment with over 700GB of memory and 800+ CPU cores, eliminating garbage collection

---

[22] ComputerWorld Profile: Tony Bishop, July 9, 2007

[23] The utility computing payoff
By virtualizing its Java application infrastructure, Wachovia Bank achieves a 300%-plus ROI.
By Julie Bort, Network World, 09/25/06

[24] A New Foundation: SOA Implementation for Business Transformation
By Nancy Feig, Wall Street & Technology, April 26, 2007

pauses and enhancing throughput. Targeted client-facing applications in fixed income and capital markets were migrated to a NAP.[25],[26]

15. Introduce solid state disks to act as a fast caching mechanism, reducing response times.

In some cases, especially numerical processing applications that ran valuation and risk, large file sets needed to be accessed very fast in their native format. Solid state disks let the application see the files as disk files but the transfer rate of 200 Mbytes/second is twice as fast as the transactional cache, offering high-speed read/write access and persistence integrity. This was especially helpful for certain long running image processing applications that consumed considerable resources and time on a regular basis.[27]

16. Introduce grid brokers for targeted applications with long running calculation requirements.

Some applications have long running units of work such as transformation engines, risk calculation engines, statement generation and PDF creation. These applications would typically benefit from a finer grained control of resources to better enable concurrent processing and take advantage of specialized processors as needed. A grid broker enables such application type classes to run more effectively on the same resources, because it manages requests for calculation services based on a real-time view of demand and supply. In some cases the applications require some restructuring. At Wachovia, grid brokers were used to streamline processing in PDF generation for certain International business operations, reducing the overall time to process by 50% on half the hardware. For risk processing, a projected $6 million in hardware expenses was reduced to $1 million through the use of grid brokers on commoditized blade racks.[28],[29]

---

[25] Wall Street-Style Power, By Penny Crosman,
Wall Street & Technology, InformationWeek,
Apr 14, 2007 12:00 AM (From the April 16, 2007 issue)

[26] Split-Second Securities Trading at Wachovia, May 21, 2007 (Computerworld)

[27] A New Foundation: SOA Implementation for Business Transformation
By Nancy Feig, Wall Street & Technology, April 26, 2007

[28] The utility computing payoff
By virtualizing its Java application infrastructure, Wachovia Bank achieves a 300%-plus ROI.
By Julie Bort, Network World, 09/25/06

[29] A New Foundation: SOA Implementation for Business Transformation
By Nancy Feig, Wall Street & Technology, April 26, 2007

17. Introduce a mediation framework, based on network attached appliances that manage authorization and authentication in a strategic manner at wire speed, while providing a protocol converter that facilitates inter-application communication through web services.

    The introduction of a specialized Network Attached Appliance (NAA) that performed as a protocol gateway greatly simplified security enforcement at Wachovia because it inspected each message at wire speed and performed appropriate security checks based on policies while converting formats at wire speed. NAA helped streamline the integration of online mortgage processing at Wachovia.[30]

18. Create the first PEDs for specific application groups based on line of business, operating characteristics and degree of collaboration between target applications. Run them on unique vertically cooled container racks.

    The integrated frameworks and ensembles provided the first PEDs to run fixed income and loan processing applications that directly affected revenue.[31]

Operate

19. Establish an on-boarding process and SWAT team so that application and operational teams can adjust to the new RTI paradigm, thereby ensuring a smooth transition.

    The new paradigm CIB engaged in broke many old operating rules including ownership of infrastructure and the comfort of owning all aspects of an application. It was easy to misuse the new infrastructure and then claim it did not work (e.g. using a grid broker where a service broker would make more sense, using an NAP for single-threaded applications and claiming it did not perform). The on boarding team profiled every candidate application ahead of deployment, performed an architectural analysis to pick the optimal platform and helped the team migrate to the platform, helping rewrite the operational procedures. This on boarding team was able to migrate 50 applications in the time it would normally have taken for 10.[32]

20. Establish infrastructure services that make RTI enhancements easier to consume.

    As part of the alignment exercises, CIB management identified key services that needed to be created to end the proliferation of point solutions (e.g. risk service, position service, calendar service). Instead, these typical point solutions became Business Support Services (BSS). These

[30] Split-Second Securities Trading at Wachovia, May 21, 2007 (Computerworld)

[31] Waters, TRADING FLOOR, Certoma Floors It

[32] NETWORKS, For the Greater Grid, (Waters), By Bob Giffords

services needed to be operationalized in a dynamic demand-based infrastructure that would be able to meet the anticipated service requests that this new operating paradigm would create. The infrastructure services needed to support these business services and included the ability to access more CPU, certain types of CPU and memory in a virtualized manner. The maturity of the ensembles and its interactions with mediation services made Information as a Service (IaaS) possible, which in turn makes BSS feasible.[33]

21. Establish IaaS by providing a federated model of key enterprise data entities that will reduce large-scale file movement for decision support and minimize the proliferation of file copies that occur to support reporting.

    IaaS was a critical CIB initiative due to the proliferation of databases and unstructured data, which resulted in excess network traffic caused by overlapping requests for data. Part of this initiative was establishing a golden record project across CIB, so that ownership and grouping could be properly established. It then became apparent that tools focused on structured and unstructured data identification were needed, along with a platform that enabled federated views, reducing data movement. The initial implementations in client management and for internal IT usage for MIS led to 25% in cost avoidance to what is normally a given in traditional environments.[34],[35]

22. Establish the self healing operating environment by introducing tools that perform real-time event correlation on events, and then predict failures.

    As the CIB environment grew, it become more complex initially. New infrastructure mixed with old, new applications being introduced at a prodigious rate. Failures were bound to happen as new processes were put in place to manage the dynamic infrastructure, even as the older infrastructure was maintained. Established networks were stressed and strained because they weren't designed to handle the explosive growth. Predictive tools were required that would first alert operations of upcoming problems, so that operations could have sufficient warning and make adjustments.[36],[37]

---

[33] Hands-On SOA Governance Practices and Processes
Information World SOA Executive Forum (5/16/2006), Sheppard Narkier

[34] Dealing with Technology, Wachovia Tailors Grid for Growth

[35] Waters, Certoma Floors It

[36] TechWorld, Virtualization tools balance data center loads,
Katherine Walsh, June 10, 2007

[37] Waters, NETWORKS, For the Greater Grid

Sustain

23. Establish portfolio management of applications so that migrations of the business value chain can be managed and new infrastructure services can be tracked.

    CIB started this process early in the align phase where all applications by LOB were identified in terms of their business value chain metric (i.e. business functions they covered), and the products they covered. This was done to aid in the infrastructure migration plan, but it was also used to identify functional overlap and functional gaps. Portfolio management is the application of an investment discipline to all IT project decisions. It breaks siloed mentalities where only the concerns of a single project of LOBs dominate the decision-making process. CIB used this process to perform quarterly progress reviews across all LOB projects. A repository was created to capture the salient features of each application, service and component, and the user demographics associated with them. It was used as a bottom-line assessment tool to show the business where resources were spent. Portfolio management is a key building block of the governance process that will be discussed in a later section. [38],[39],[40]

24. Create metrics dashboard for multi-stakeholder consumption so that ROI discussions about IT can be grounded in fact.

    The CIB portfolio management process needed hard numbers to assess the performance of new investments and legacy implementations. This helped guide investments for new initiatives and show how investments changed the Quality of Experience profile for key application sets. In order to do this, all performance tool outputs needed to be correlated so that trends could be shown regarding the total cost of a transaction. There were many perspectives to be captured: service level adherence, consumption by infrastructure class, chargeback projections, etc. The purpose of the dashboard was to provide an active feedback mechanism that would augment portfolio management for executives so that their strategy could be grounded in execution reality. This dashboard changed initial assumptions about priorities; in particular, the amount of upgrades to the base infrastructure that would be required was underestimated.[41],[42]

---

[38] SCREEN, Creating Virtualization

[39] Wall Street Technology, Governance From the Top

[40] Align Online, Embracing Change

[41] Waters, NETWORKS, For the Greater Grid

[42] eWeek.com, Tech Earns Interest at Wachovia

25. Establish product management that will facilitate the introduction of other RTI products through an integration strategy based on established investment principles.

    The reuse of design and services or components became a critical part of the CIB strategy, because there were many applications types that have performance characteristics exceeding the typical COTS capabilities. Too often in CIB, specialized components could not be leveraged across other projects, LOBs or the CIB enterprise. Product management is the discipline that ensures releases of the PEDs, ensembles are integrated, documented, tested and ensured to work at certain levels of performance. Product management ensures adherence to product principles about what functionality is included and who will own the components if harvested. It is about providing an internal software vendor service for strategic choices of functionality. CIB established product strategies for the ensembles and their frameworks. CIB defined an internal product harvesting process to find and evaluate candidates to add to the service. This ensured that innovation was sustained at the infrastructure and application level.[43],[44],[45]

Governance: The Key to Sustaining RTI Value

The introduction of Fit-for-Purpose RTI as an approach for IT to become more agile is not just about buying particular sets of technology. It is about a shift in IT thinking that will give IT the grounding required to become a strategic enabler of business goals as opposed to being just a cost center. IT and business need to work together and think of IT as an investment. This requires a shift in process, people skills and technology usage. This shift needs to have an underlying theme about investment discipline which requires IT Governance.

IT Governance has been a controversial topic for some time, because while everyone agrees that governance is critical to the long-term success of Enterprise IT, there is rarely a consensus about how it should be done. During the boom years of 1996–2000, any kind of rationalization was ignored because the internet was so disruptive that reaction to service was the key strategy. After the crash of 2000, the result of this non directed proliferation of applications became evident as the amount of customer facing applications and product databases routinely tripled across large enterprises. Suddenly Governance was put established to essentially stop proliferation, but it did not address renationalization as budgets were tight and the focus was on critical applications that kept the business running. The movement away from large SMP machines to commoditized blade racks after the crash was prompted

[43] Banks Take Their Places on the Grid, By Paul Allen; Photography James Leynse, Wall Street & Technology, May 15, 2006

[44] TechWorld, Virtualization tools balance data center loads, by Katherine Walsh, June 10, 2007

[45] Align Online, Case Study: SOA Delivers for Wachovia – With Interest

by the belief that significant cost reductions would ensue, but this movement caused a different set of proliferation issues, including excess complexity and server sprawl which has been covered in this document.

In neither case were the issues of proliferation or rationalization addressed, and with the resurgence in the economy occurred in 2004, the consequences of not having a coherent governance strategy began to emerge in terms of runaway costs, unnecessary complexity and reduced agility.

IT Governance cannot be a passive rubber stamp, nor can it be a stifling, irrelevant ivory tower exercise. IT Governance is about incorporating the discipline of IT Portfolio Management, along with a set of sustainment processes for product management, incentive alignment, skills augmentation and the creation of an internal IT marketplace that both reuses and produces IT assets. This marketplace requires the currency of *trust* as its catalyst and *trust* can only occur when accurate and timely information is readily available at the appropriate level of detail for each stakeholder group.

Portfolio Management

Portfolio Management is about making sure that investments are guided as they relate to business goals. The guiding principle for IT Portfolio Management is that every dollar spent on IT is an investment. IT competes for investment dollars in an organization with every other department (sales, marketing and client services). When a business makes an investment in IT the business expects some or all of the following.

- Better ability at revenue generation, better efficiency, or risk reduction.
- Greater agility so that business can respond to market and regulatory pressures faster.
- Insight into the alignment of IT and the achievement of business goals.
- Can IT show alignment and support to business goals?
- Can impact be measured, is value traceable across stakeholders.

Portfolio Management Vision

Investment in IT can be managed as a portfolio of assets that have cost and value over time. Portfolio management, which includes governance, must put the assets into aligned context with the business goals. Business and IT must take an investment management view of these assets so that the collective senior leadership understands, guides, realizes and governs the portfolio. In this new paradigm IT is about showing its relevance.

Components of the Vision Include:

- Aligned business and IT that improves business strategy results.
- Communication between business and IT is significantly improved.
- Business values of delivered assets are measurable and tangible.
    - Other intangibles can be viewed as well: customer satisfaction, efficiency, and regulatory compliance.
- Risks in technology can be measured in terms of skills, product maturity, etc. The risk of doing nothing can also be measured.
- Strategic cost avoidance can be practiced because there is structured analysis of duplicate efforts.

- Investment discipline is exercised so that low-value efforts are identified early and stopped.
- There is portfolio focus so that uncoordinated funding is reduced, emphasizing shared services where possible.
- Investment leverage is encouraged.
    - Cross-LOB sharing is measured and tracked as a key to rationalization and efficiency.
    - Contribution to IT assets and compliance to standards are rewarded according to value.
    - Optimization and agility is fostered through the use of standard best practices.
    - The results can be shown to leverage investments in the RTI environment.

Managing RTI through Portfolio Management

Migrating to RTI efficiently requires governance of existing services and assets and a clear understanding of future requirements. IT executives not only require an inventory, but also an integrated view of the information that explains those assets. A repository is required to house all of the portfolio management information, which in turn is collected from numerous sources in a large enterprise, including CMDBs, various inventories, performance statistics repositories, etc. The IT assets have many levels of detail, but executives think of IT assets in terms of services, applications and infrastructure. Executives must be able to access knowledge that enables them to understand their asset base so that they can provide directives for change and then have architects, project managers, developers and operations all be able to have a tangible connection to those targeted assets, while accessing greater levels of detail to make the changes.

The Key Management Principles for IT Governance

- Business and IT executive buy-in is crucial with IT as an investment as the guiding principle: without this the initiatives which are so transformative will have no grounding. Executives must actively guide the investments; grounded in business capability, risk tolerance and the need for agility.
- Organizational and incentive structure alignment is critical to success. Executive management must be paid based on an ability to hit RTI goals that include supporting specific IT governance goals. The problems of domain of control need to be addressed. Most executives feel very uncomfortable if they do not own all assets that their applications run on. Shared services, frameworks and shared infrastructure break that domain of control, so organizational structure, processes and incentives must be created that will accommodate this new paradigm.
- Dynamic governance is critical. This requires that applications and infrastructure in transition must be accounted for. Typically infrastructure, procedures and applications become more complex during the transition before they get simpler. Exceptions need to be considered and accounted for. Change needs to be embraced, and that includes changes to applications that are not necessarily strategic, but are required to run the business currently. It is during this time of change that a new set of shared values about how to use and operate a Fit-for-Purpose need to be defined and promoted.
- Fit-for-Purpose RTI must be supported as if it were a utility with processes to create and review SLAs and create tangible chargeback models.

- Feedback must be part of this process which includes performance and availability measurement, goals compliance and operational efficiency. Executives must have transparent access to information so that they can make informed decisions.
- Training must be targeted to make the change to a dynamic RTI and a dynamic management process. Training must be thought of as a maturity process requiring levels of training for all stakeholders.
- An IT marketplace should be established so that asset designs, working products, asset reviews and hard performance data can all be viewed as part of the IT portfolio management repository. This repository will enable the market place dynamics that can effectively take advantage of the Fit-for-Purpose RTI platform, fostering agility, transparency and cost effectiveness.

The goal of these management principles is to enable the organization to be agile so that it can effectively take advantage of the resources and better meet the goals of the organization.

Wachovia CIB Experiences with IT Governance

Wachovia CIB undertook multiple threads of action to restructure IT governance. The following key threads are explained.

- An Architecture group was reformulated to focus on enabling the design of products, aiding teams undertaking new projects and providing objective standards that could be used to evaluate designs and systems. This architecture team set standards for component harvesting, framework design and ensemble creation and engineering. This group also conducted strategy reviews that highlighted what services needed to be consolidated or created.
- A Portfolio Management group was formed to facilitate the information flow between executives and asset creation staff. This group built a variety of tools that gave insight into the overall progress of the portfolio management strategy, leveraging a repository to generate monthly reports on ideal state migration and contributions to the asset base and user demographics for all applications in CIB. This group worked closely with various operations and performance measurement groups in order to build a comprehensive view of the asset base operating profiles.
- A number of framework groups were created to channel investments in reusable service architectures. These groups built a client access framework, a server framework and a mediation framework. All of these frameworks had multiple clients, harvested existing designs and programs where feasible (in addition to creating new functionality). An emphasis was placed on integrating new Fit-for-Purpose RTI technologies that would reduce cost, increase efficiency, stop datacenter sprawl and promote agility and increase throughput and performance.
- The framework teams took direction from a technology steering committee comprised of senior members of all stakeholder groups. These members were aligned with specific LOBs and thus were able to articulate real needs. Executives were kept informed monthly of all investment.

This forced a product management view of development. If the client did not ask for it, or approve a feature submitted for review, no investment would be made. It created a reliable integration strategy that consistent delivered functionality in timely increments.

- Quarterly strategy reviews were held by the CIO to determine progress to plan on all investment tracks. Architecture and Portfolio Management groups arranged the material for these reviews. These reviews fed the strategy process for the CIO and direct reports.
- Monthly user demographic reports were produced for all applications in all LOBs to understand changes in demand and its potential effects on the strategy.
- A developer university internal site was created as a repository for best practices, changes to technology, how to's on RTI, asset contribution, architecture reviews, etc.
- Incentives were structured from CIO directs on down to align with use of the asset base and contributions to RTI.

---

[i] Quality of Experience (QoE): A maturity model premised on the belief that cost efficiency and service quality must be balanced in order for any technology approach to be sustainable, and have continued business buy-in.