White Paper

Cloud Computing for Financial Markets

Executive Summary **3**
- Business Drivers / Strategy 3
- Key Performance Indicator Improvements 3
- Execution Strategy 3

Cloud Computing Overview **3**

Enterprise Cloud Design **4**

Use Case Examples **5**
- The Desktop Experience, Revenue Generating Use Case 6
  - Day in the Life of a High-Performance Trader 6
  - Requirements to Fulfill the Trader's Need 6
  - Problems and Consequences in Meeting that Need 6
- The Desktop Experience, Client Facing Use Case 9
  - Day in the Life of a Client Account Manager in the Retail Brokerage 9
  - Requirements to Fulfill the Manager's Need 9
  - Problems and Consequences in Meeting that Need 9
- Decision Support Infrastructure Use Case 12
  - Day in the Life of Decision Support Systems 12
  - Requirements to Fulfill Decision Support Systems' Needs 12
  - Problems and Consequences in Meeting that Need 12
- Application Delivery Use Case 15
  - Day in the Life of the Typical Employee in a Retail Brokerage Branch 15
  - Requirements to Fulfill the Retail Branch's Needs 15
  - Problems and Consequences in Meeting that Need 15
- Work Shift Adjustment Use Case 17
  - Day in the Life of the Application Suites in an Enterprise 17
  - Requirements to Fulfill Decision Support System's Needs 17
  - Problems and Consequences in Meeting that Need 18
- Disaster Recovery Infrastructure Use Case 21
  - Day in the Life of Disaster Recovery Platforms 21
  - Requirements to Fulfill Disaster Recovery System's Needs 21
  - Problems and Consequences in Meeting that Need 21

Design Tips **25**

Cisco Cloud Enablement Strategy **27**
- Workload Mobility 27
- Security and Control 27
- Integration and Interoperability 27

Next Step: Contact Cisco **28**

About the Authors **28**

Executive Summary

Enterprise IT is changing in the way that applications, information, content, compute, storage, and network is deployed and consumed. The move is toward the service delivery of IT through cloud computing, a dynamic and service-oriented infrastructure that allocates infrastructure supply to business demand as needed, just like a utility.

The key design consideration of cloud computing is the abstraction of underlying applications, information, content, and resources. Physical resources are housed in massively scalable, tailored resource pools configured for dynamic provisioning of virtualized resources. This abstraction makes the underlying resources easier to manage and provides the basis for more effective management of the applications themselves. The cloud operating environment is more agile, reliable, and efficient; and because it is abstracted, it can evolve as new technologies emerge while avoiding many negative impacts that infrastructure changes normally cause. A cloud of computing resources can exist inside or outside a corporation’s data center, which enables migrating nonstrategic services to an application service provider without sacrificing service levels or cost savings.

Business Drivers / Strategy

- **Simplify Delivery of IT**: Deploying infrastructure and applications often takes weeks or months.
- **User Experience**: The business has been promised improved user experiences that have not materialized, which leads to missed revenue and productivity opportunities.
- **Agility**: Beyond speeding delivery of supply, the business needs to dynamically scale up or down in real time to have resources where and when needed.
- **Environment Efficiency**: Large portions of the IT infrastructure remain poorly utilized.

Key Performance Indicator Improvements

- **Time to Market**: Reduction by as much as 50 percent.
- **Total Cost of Ownership**: Reduction in capital and operating expense by up to 60 percent
- **Performance**: Improvement by as much as 50 percent.
- **Data Center Productivity**: Reduction in packaging, deploying, and changing applications and infrastructure by as much as 70 percent.

Execution Strategy

- **Virtual and Dynamic**: Allocate low-overhead virtual resources from physical ones dynamically and in real time.
- **Reduce Resource Costs**: Once deployed into a cloud infrastructure, resources can be internal or external to the organization depending on what is most cost effective.
- **Robustness and Performance**: With virtual and dynamically allocated resources, replacing failed or scaling up resources as needed becomes significantly easier.

Cloud Computing Overview

Cloud computing is often viewed as a revolutionary, disruptive innovation. However, considering the use of grids and the notion of IT utility, it is clearly a natural evolution of the digital supply chain. The concepts of cloud computing—networked infrastructure resources, available for a price and accessible over a standard interface—are quite similar to preceding ones, with the primary differentiator being that these resources are now supplied from outside the enterprise firewall.

Although today cloud computing is widely regarded as a consumer or small business platform, it will play an increasingly important role in the data center of the future, particularly as resource demands continue to grow for applications such as real-time collaboration. Enterprises raising the efficiency of their internal IT utility to higher levels will purposely constrict their internal capacity available, perhaps shifting from engineering the utility capacity to meet 99 percent of average peak capacity demands to 95 percent. When workload demand exceeds available capacity a choice must be made: throttle or suspend lower priority workloads, or meet the demand with supplemental capacity outside of the enterprise. Whether that demand spike occurs due to an outage, an unforeseen demand surge, or simply because the utility operators wish to drive higher efficiencies, cloud computing is well suited to meet that demand. In fact, as the technologies underpinning cloud computing mature, progressive organizations may permanently shift a growing proportion of infrastructure capacity to the cloud and eventually migrate most if not all of it. Assuming enterprise security concerns are addressed, such decisions will be made based on economics rather than technology.

The need for skilled IT executives will not go away during this process. The lessons they have learned in resource sharing from grids, and the business skills and control systems for dynamic application service management in an IT utility, are essential to exploiting the benefits of cloud computing. Enterprises that master these skills early will have a distinct advantage.

Enterprise Cloud Design

The basic cloud is composed of pooled resources (compute, network, and disk) that are dynamically provisioned to meet the demands requested of it. Real-time and on-demand virtualization technologies are used heavily for rapid deployment of resources to meet the demand. In general, three primary services are provided by the cloud infrastructure, though they are each interrelated:

- **Infrastructure as a Service (IaaS)**: Raw compute, network, and disk resources on demand. Often virtualization technologies help in real-time reconfiguration of environments.
- **Platform as a Service (PaaS)**: A complete computing platform available on demand for design, development, testing, deployment, and hosting of applications. This includes everything provided by IaaS plus the development environment, application runtime, middleware, caching services, security services, integration services, and so on.
- **Software as a Service (SaaS)**: A server-based version of software application or content channel that a remote thin client can connect to and operate with. Management of the software is centrally controlled and preferably is managed as virtual instances available on demand and typically built upon a PaaS.

A cloud can be operated by your organization (internal) or a third party (external). There are various tradeoffs to creating internal or buying external cloud resources. Deciding on the level of service (platform, software or infrastructure) will affect the level of investment in infrastructure and personnel. The strategic importance of an application, combined with its performance parameters, will dictate whether it is better to outsource an application externally or keep it internal. See Table 1 Cloud Decision Model for details.

Table 1. Cloud Decision Model

Typical Problem	IaaS / PaaS / SaaS	Ensemble	Internal or External Cloud
Need extra CPU and memory for spikes of trading during a volatile market for a few hours at a time to run high-risk analytics.	IaaS: From an extended pool to minimize security and data transfer issues.	Numerical Processing Ensemble	Internal cloud: Preferred for providing quick (within minutes) resource allocation.
Need extended platform resources available to print end-of-month statements. Don't want to invest in more infrastructure, and current data center sharing schemes cannot guarantee service window times.	PaaS: Provides a tailored footprint for the processing required, including sufficient disk, CPU, memory, and bandwidth. Would need encryption for large bulk data transfers.	Numerical Processing Ensemble with heavy emphasis on large data movements.	External cloud: Operations can be run remotely by the renter. If more resources are needed to meet time windows, this would be covered under the service level agreement in terms of incremental cost to supply above and beyond.
Manage costs for GL and Oracle financials internally.	PaaS: Generate allocated chargebacks for resource used in apportioning costs to process all feeds from federated LOB.	Information Integration Ensemble combined with Numerical Processing Ensemble	Internal cloud: Used as a transparency measure to get LOBs to streamline their input. Too often GLs must process hundreds of input feeds every night. Internal PaaS lets IT treat this processing as a business.
Stop building internal support staff to house Taleo, HR, timesheets, and so on.	SaaS: Get a tight services contract so that these become externally provided SaaS.	Complex Transaction Ensemble	Internal cloud: Must resolve issues of security, possibly two-factor authentication .
Reduce investments in point solutions for common services such as Date/Holiday service position service.	SaaS: This should be created with contracts in mind.		Internal cloud: To minimize delay.

Use Case Examples

The following six use cases are common in organizations that are significantly challenged by traditional solutions. Each one provides an analysis of the challenges, followed by how they benefit from the availability of cloud infrastructure.

- **The Desktop Experience, Revenue Generating**: Centralization of compute processes away from the desktop and onto the cloud to improve the trader’s experience and simplify management.
- **The Desktop Experience, Client Facing**: Replacement of thick-client technologies with thin-client/SaaS solutions to improve cost and simplify management.
- **Decision Support Infrastructure**: Transformation of fixed and discrete grids to a more efficient and effective grid utility.
- **Application Delivery**: Transformation of fixed and discrete application resources into a more flexible SaaS solution.
- **Work Shift Adjustment**: Transformation of a fixed infrastructure into a time-of-day and/or business-event-driven dynamic cloud infrastructure.
- **Disaster Recovery Infrastructure**: Transformation of idle resources into a PoC, performance testing, UAT, and development environment utility, with faster recovery time in the case of a disaster.

Each use case has a brief “day in the life” description of the challenges faced by a specific aspect of a financial institution. This is followed by a description of requirements necessary to meet the need, and a section of the problems that typically manifest themselves when traditional approaches are applied.

Diagrams depict a high-level view of the environment discussed and are tagged with problem areas and followed up with a solution diagram. Each diagram has areas of interest tagged with numbers, which are related between the diagrams (problem in red and solution in green).

Cisco has developed a family of products and partnerships that address the many challenges of deploying and operating a cloud environment. These use cases are meant to highlight the general approaches and how a Cisco oriented cloud infrastructure can enable the environment and simplify the architecture.

The Desktop Experience, Revenue Generating Use Case

Day in the Life of a High-Performance Trader

A trader in financial services must make critical decisions quickly because the market moves at sub second speed. The trader must process enormous volumes of information and make judgments expeditiously to stay ahead of the market. If a trader is not kept informed in a timely manner, then the firm could miss an opportunity or take a considerable loss.

Requirements to Fulfill the Trader's Need

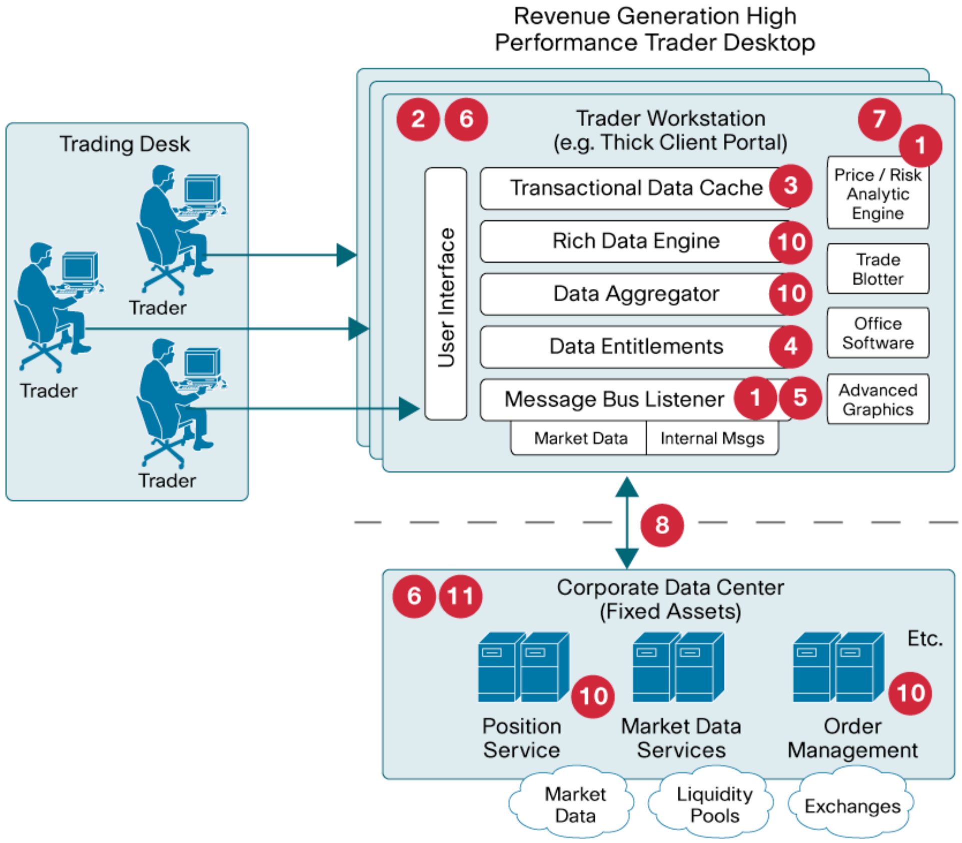
A trader needs information from the markets delivered in a timely manner; incoming message rates could be over 1000 messages per second from multiple markets. A trader needs to run models that accept the influx of real-time messages while calculating strategies on the fly. A trader also needs sophisticated visualization tools that make it easy to see predefined alerts and trends that can be acted upon. A trader needs streaming news sent to the desktop so that information remains current. If a trader is unable to make a decision due to lack of information, unreliable operation of the models, or slow alerting, it could cost the firm millions of dollars; a trader's desktop and its entire supporting infrastructure must be ultra-reliable.

Problems and Consequences in Meeting that Need

Trader desktops are one of the most extreme examples of users that demand large "personal" infrastructures that can be tailored to an individual's taste. The consequences are:

- **An Overwhelmed Desktop Affects User Experience**: Platforms must process large volumes with low latency (under 40 ms per transaction), handle thousands of inputs per second, render real-time graphics, run real-time alerts, keep a large cache of highly volatile data, and be ultra reliable as downtime can be measured in millions of dollars lost per hour. The trend for some time has been to expand the workstation to meet the demands of the individual trader. This is costly, and ultimately does not solve the problem as the pace and complexity of growth outraces workstation capacity.
- **Desktop Workstation Sprawl**: Increased cost and operational complexity; some solutions place multiple workstations for a trader. This adds cost, increases complexity and causes reliability problems. It can overtax the network as it serves multiple workstations per user.
- **Collaboration**: Traditional applications treat collaboration as an afterthought if at all; however, this is a source of major product and service differentiation. Missed business opportunities or clunky interfaces are the result.
- **Operational Complexity Decreases Reliability**: Each trader tends to have their own configuration, and the degree of dedicated personnel and other support costs assigned to a trading desk causes costs to be out of control. Problem determination is harder, making it harder to maintain reliability.
- **Agility Suffers**: New features are hard to roll out; the whole purpose of giving traders large workstations is to maximize agility, but the fast moving market pressures that force sprawl also make the trader platform less stable, impeding the migration of application changes. In the end, the primary drivers of agility and performance cannot be met successfully.

Figure 1. Traditional Trader Desktop



Problem Areas

1 **Workstation Performance**: Peak market data traffic competes for CPU and I/O resources with longer running analytic calculations Scale by increasing desktop capacity; adding CPUs to separated processing promotes uncontrolled growth.

2 **Time to Market**: Deployment of new functionality must be rolled out rapidly to the Traders. Each Trader has their own configuration; hundreds of individual changes required for upgrades; inhibits agility; causes operational instability.

3 **User experience**: Large amount of cache required for large influx of market data, streaming data, the data is very volatile. Non-optimized infrastructure causes frequent cache updates and garbage collection pauses, disk utilization runs high.

4 **Security**: Access to resources must be secured. Exposure to organization is high from market data providers. Security and permission measures are often ad-hoc, and manually checked only periodically

5 **Workstation Scaling**: Market Data rates are constantly increasing. Processing and network load stress individual trader workstations, forcing upgrades.

6 **Shared Infrastructure**: Infrastructure should be shared for common services. Caching and analytic engines are deployed per workstation and datacenter resources are fixed causing sprawl, escalating costs and operational complexity.

7 **Collaboration:** Interactive collaboration is needed between users to share information. Traditional applications typically lack collaboration features, such as IM and sharing a view.

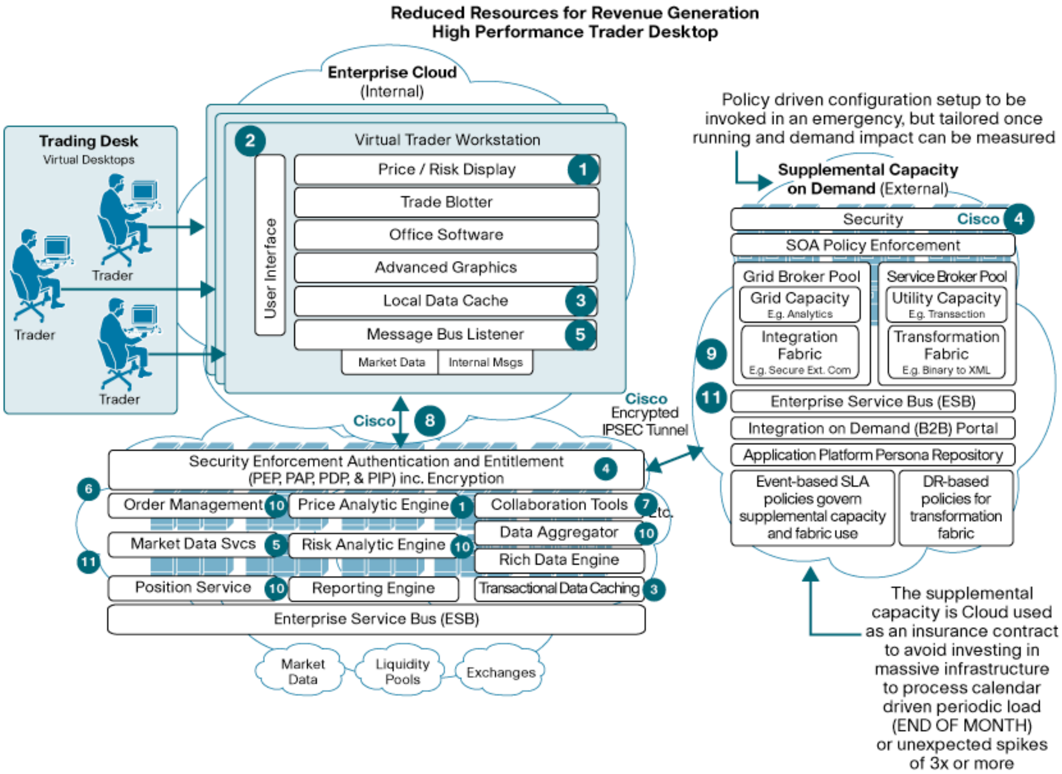
8 **QoS**: Market data messages, such as trades and orders must be top priority. All network traffic has equal importance; trades and orders compete for bandwidth.

9 **Load Spikes**: There are spikes 2x-3x baseline peak that infrequently occur. Invest in additional and costly infrastructure or allow QoS to suffer.

10 **Performance**: Client facing Java applications must perform consistently. Java applications suffer from garbage collection pauses and JVM memory limits, which can greatly affect reputation.

11 **Physical Connectivity:** Density of servers continues to increase. Each server has multiple connections to the network and SAN, which raises costs, lowers reliability, and makes maintenance more difficult.

Figure 2. Cloud Solution for Trader Desktop



Solutions

1 3 Performance: Local caching requirements reduced as most volatile persistence is processed and held by the Cloud, which has tailored caching engines that streamline processing. Increased market data activity is processed by the cloud as it transparently adds resources to meet demand.

2 Time to Market: Deployment of new functionality must be rolled out rapidly to the Traders. Enterprise cloud provides consistent service builds for common functions; upgrades occur seamlessly.

4 Security: Access to resources must be secured. Cisco Policy Manager and Policy Enforcement Points; ensures WIRE SPEED fine grained auth disallowing invalid requests before they waste resources. Business Policy Driven: (Client Trader Entitlements)

- Policy Enforcement Points (PEPs)
- Policy Administration Points (PAPs)
- Policy Decision Points (PEPs)

5 Workstation Scaling: Market data rates are constantly increasing. CPU intensive processing demand, volatile data management and disk management stress removed from Trader Server Platform, reducing the need to add resources per trader continuously.

6 Shared Infrastructure: Infrastructure should be shared for common services. Trading desks share infrastructure, such as caching or analytics engines; reduce datacenter sprawl, simplified cabling and escalating costs.

7 Collaboration: Interactive collaboration is needed between users to share information. Application integration with Cisco Unified Communication Manager.

8 QoS: Market data messages, such as trades and orders must be top priority. Network traffic for trades and orders transaction types are prioritized by Cisco's QoS Policy Manager.

9 Load Spikes: There are spikes 2x-3x baseline peak that infrequently occur. Dynamically tap supplemental third party cloud capacity to meet spikes.

10 Performance: Client facing Java applications must perform consistently. Supplemental Java processing with a network attached JVM processor (Azul) which eliminates garbage collection pauses, accelerates inter-communications with <40ms response time and supplies 864 JVM cores / 768GB memory.

11 Physical Connectivity: Density of servers continues to increase. Cisco Unified I/O combines network and SAN connectivity into a single converged network adapter connected to a Cisco Nexus switch to run FCoE. Quality and Performance over FCoE is enabled via Cisco's Priority Flow Control (PFC) providing a lower cost environment with higher availability that is easier to maintain.

The Desktop Experience, Client Facing Use Case

Day in the Life of a Client Account Manager in the Retail Brokerage

Retail brokerages are the face of financial enterprise, providing differentiated services to clients and fostering client loyalty and retention. Retail brokerages strive to provide a comprehensive set of client services and expect them to be easily navigated to instill confidence in the client. During heavy market activity it is essential that a client account manager be able to counsel a client by utilizing timely market data.

Requirements to Fulfill the Manager’s Need

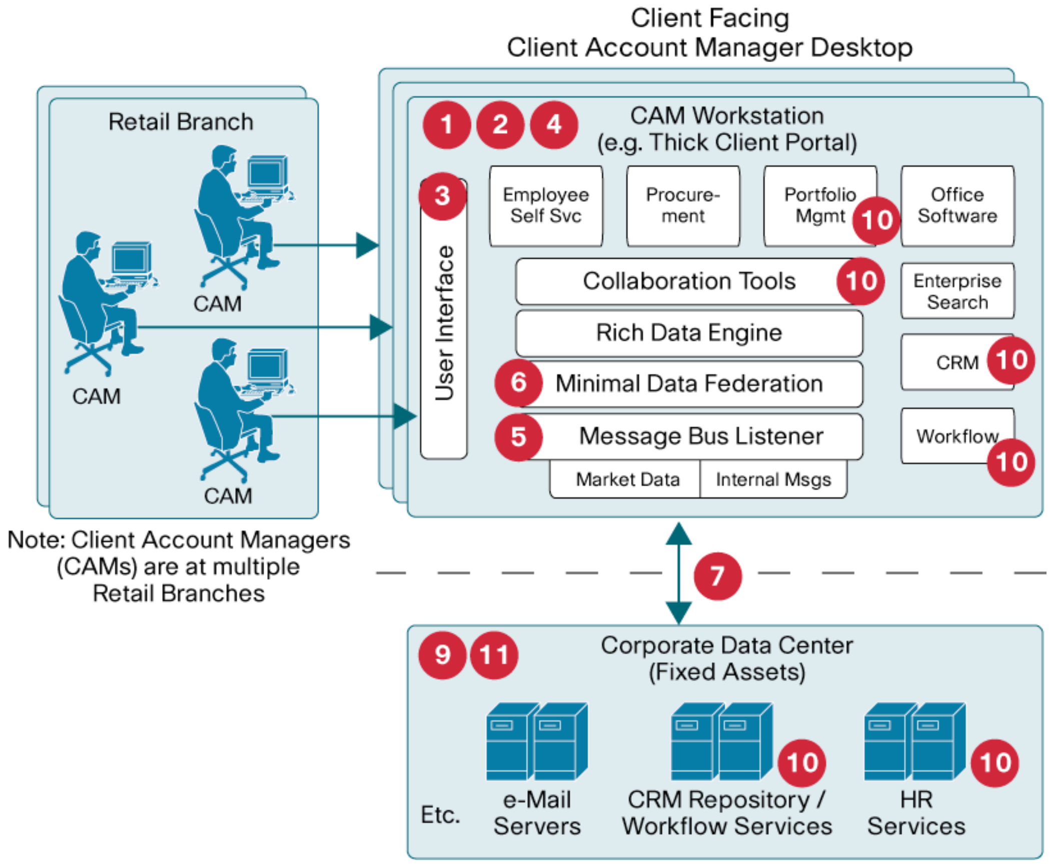
Client account managers should have streaming data and be informed of trends, use PDAs to keep current when visiting a client, and be able to procure a complete view of all accounts a client has (family members, IRA, regular investments, loans, and so on). They need to be able to run what-if scenarios for clients and quickly make presentations including projections and trends. They need easy access to client preferences, a priority to ensure that recommendations are in line with the client’s objectives.

Problems and Consequences in Meeting that Need

The typical IT answer to a retail branch’s needs is to roll out suites of applications to robust workstations that communicate with a branch server, which keeps a subset of data local to that branch. That branch server then communicates with a central hub back at a robust data center. The consequences are:

- **Slow Application Updates**: Application update rollouts are time consuming, often taking over a year, and are compounded by the problem that making changes to these systems remotely is error prone. As applications evolve and add more features, the remote, less populated branches are left behind, rarely receiving updates in a timely fashion.
- **User Experience**: Access is cumbersome as delays become more prominent because the branch server and the infrastructure are rarely updated.
- **Collaboration**: Traditional applications treat collaboration as an afterthought, if at all; however, this is a source of major product and service differentiation. Missed business opportunities or clunky interfaces are the result.
- **Out of Date Resources**: Over a period of a year, the once robust workstations become outdated, making it difficult to provide the consistent branded user experience that is critical for the business image.

Figure 3. Traditional Client Account Manager Desktop



Problem Areas

1 **Collaboration**: Interactive collaboration is needed between users to share information. Traditional applications typically lack collaboration features, such as IM and sharing a view.

2 **Time to Market**: Deployment of new functionality must be rolled out rapidly to the CAM users. Each CAM has their own configuration; thousands of individual workstations must get changes required for upgrades; inhibits agility; many upgrade opportunities are shunned for fear of operational instability.

3 **User experience**: High performance interfaces are desired. Underlying technology platform is only upgraded every 3 to 5 years.

4 **User Experience**: Collaboration tool competes with other processing that require timely response. Stressed CPU and network resources on the workstation cause inconsistent user experience at market peaks.

5 **Performance**: Conflicts exist between large content movement and the need to have interactive web processing. Stressed CPU and network resources on the workstation cause inconsistent user experience.

6 **Security**: Access to resources must be secured. Exposure to organization is high from fragmented data storage policies.

7 **QoS**: Optimized use of bandwidth given the geographic constraints that affect service quality. Inconsistent network bandwidth to regional offices and unprioritized / uncompressed data sources.

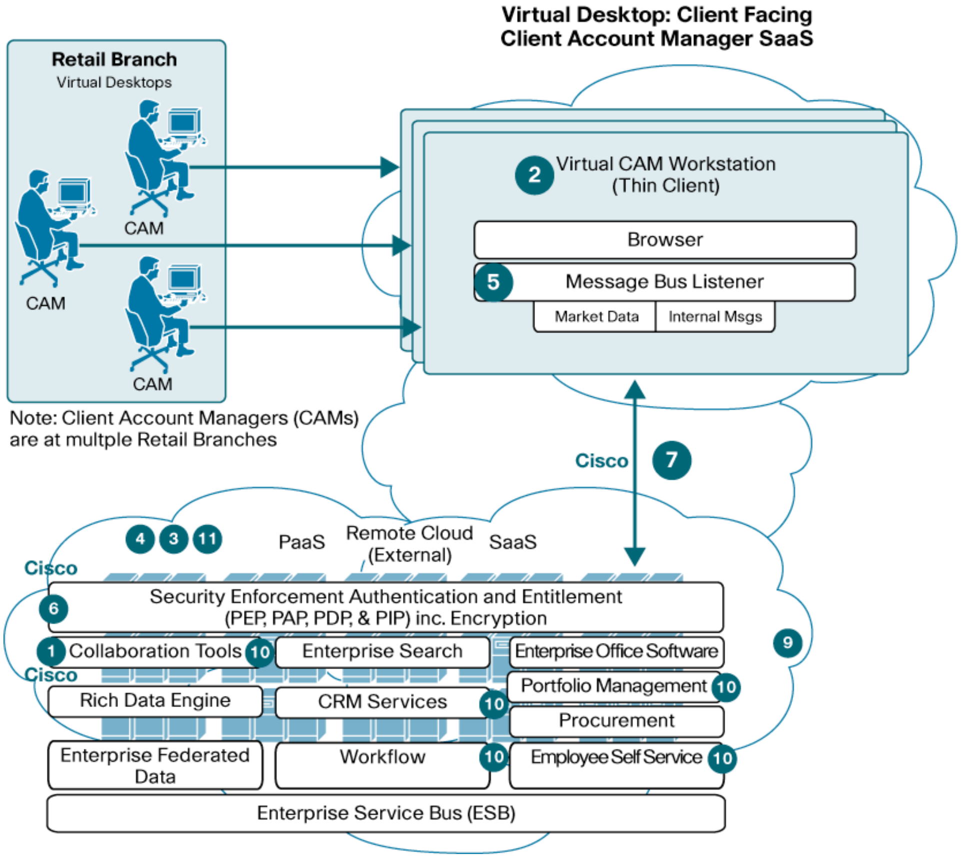
8 **Support:** The resources must be easy to maintain. The number of configurations at the remote site requires that there be staff on hand to repair and maintain the systems.

9 **Shared Infrastructure**: Infrastructure shoud be shared for common services. Caching and analytic engines are deployed per workstation causing sprawl, escalating costs and operational complexity.

10 **Performance**: Client facing Java applications must perform consistently. Java applications suffer from garbage collection pauses and JVM memory limits, which can greatly affect reputation.

11 **Physical Connectivity:** Density of servers continues to increase. Each server has multiple connections to the network and SAN, which raises costs, lowers reliability, and makes maintenance more difficult.

Figure 4. Cloud Solution for Client Account Manager SaaS



Solutions

1. **Collaboration**: Interactive collaboration is needed between users to share information. Application integration with Cisco Unified Communication Manager.

2. **Time to Market**: Deployment of new functionality must be rolled out rapidly to the CAM users. New features and changes are not hampered by workstation limitations; changes are rolled out everywhere at once: Web 2.0 infrastructure provides thick client responsiveness via a thin client.

3. 4. 5. **Performance**: Local caching requirements reduced as most volatile persistence is processed and held by the Cloud which has tailored caching engines that streamline processing. Increased market data activity is processed by the cloud as it transparently adds resources to meet demand.

6. **Security**: Access to resources must be secured.. Cisco Policy Manager and Policy Enforcement Points; ensures WIRE SPEED fine grained auth disallowing invalid requests before they waste resources.
Business Policy Driven: (Client and Trader Entitlements)
   - Policy Enforcement Points (PEPs)
   - Policy Administration Points (PAPs)
   - Policy Decision Points (PDPs)

7. **QoS**: Optimized use of bandwidth given the geographic constraints that affect service quality. Cisco QoS Policy Manager assigns priority by message type based on business rules and time schedules; Cisco CSA compresses remote traffic optimizing limited WAN bandwidth.

8. **Support**: The resources must be easy to maintain. Thin configurations at the remote site enables a standard build that can be maintained remotely.

9. **Shared Infrastructure**: Infrastructure should be shared for common services. Trading desks share infrastructure, such as caching or analytics engines; reduce datacenter sprawl, simplified cabling, and escalating costs.

10. **Performance:** Client facing Java applications must perform consistently. Supplement Java processing with a network attached JVM processor (Azul) which eliminates garbage collection pauses, accelerates inter-communication with <40ms response time and supplies 864 JVM cores / 768GB memory.

11. **Physical Connectivity:** Density of our servers continues to increase. Cisco Unified I/O combines network of SAN connectivity into a single converged network adapter connected to a Cisco Nexus switch to run FCoE. Quality and performance over FCoE is enabled via Cisco's Priority Flow Control (PFC) providing a lower cost environment with higher availability that is easier to maintain.

Decision Support Infrastructure Use Case

Day in the Life of Decision Support Systems

Decision support systems are critical as they provide layered sets of information for timely actions by executives, traders, and compliance and risk officers. These systems must aggregate data from many sources and run models to provide information in a variety of formats and degrees of detail. Decision support systems include various risk analytics, liquidity ratio calculators, general ledger, treasury projections, trading scenarios, asset allocation models, and pricing applications. Some analytics must run during the day whereas others can only run at the end of the day after all trading is done; still others must incorporate a global view and thus have multiple runs during any given day.

Requirements to Fulfill Decision Support Systems' Needs

Decision support systems must aggregate many large data sets (10 GB or greater) in multiple data formats. There is usually no more than 6 hours at night to process all of this information for one application. Typically hundreds of decision support applications run in a given night that must transform, process, and integrate data and produce summary information in that short time frame. During daytime runs there are choices a service requester can make about maximum time to calculate, number of iterations for accuracy, and so on. Daytime analytics often must accept incoming data at rates in excess of 1000 messages per second. These decision support systems often produce results in multiple formats with multiple destination points.

Problems and Consequences in Meeting that Need

The typical IT trend has been to build project grids. These have reached staggering numbers where a single project requires more than 2000 nodes to run risk analytics or other intensive processing. They cannot be used for other processing because, being designed for single projects, their configurations prevent reuse. They cannot be utilized by a variety of applications based on policy, to meet incoming demand, or to be invoked based on an event.

- **Data Center Sprawl**: Inefficient resource use drives up the cost of network, server, power, cooling, and storage and devours data center space; all the while, more resources need to be bought because at peak usage individual grids run into performance problems.
- **Network Utilization**: Proliferation of project grids leads to excess bandwidth utilization, and the stress on data center floor space causes inefficient use of the network.
- **Growth Driver**: The regulatory environment forces the need for better risk metrics and continues to force individual grids to grow, even as the apparent utilization of resources continues to dwindle below 25 percent.
- **Lack of Business Alignment**: Grids do not expand proportionately with revenue, creating an alignment mismatch that starves agility.
- **Support Costs**: The skill set for grid maintenance is hard to find and expensive.
- **Idle Capacity**: Grids set aside for disaster recovery lie dormant waiting to be utilized.

Figure 5. Traditional Decision Support Distributed Computing

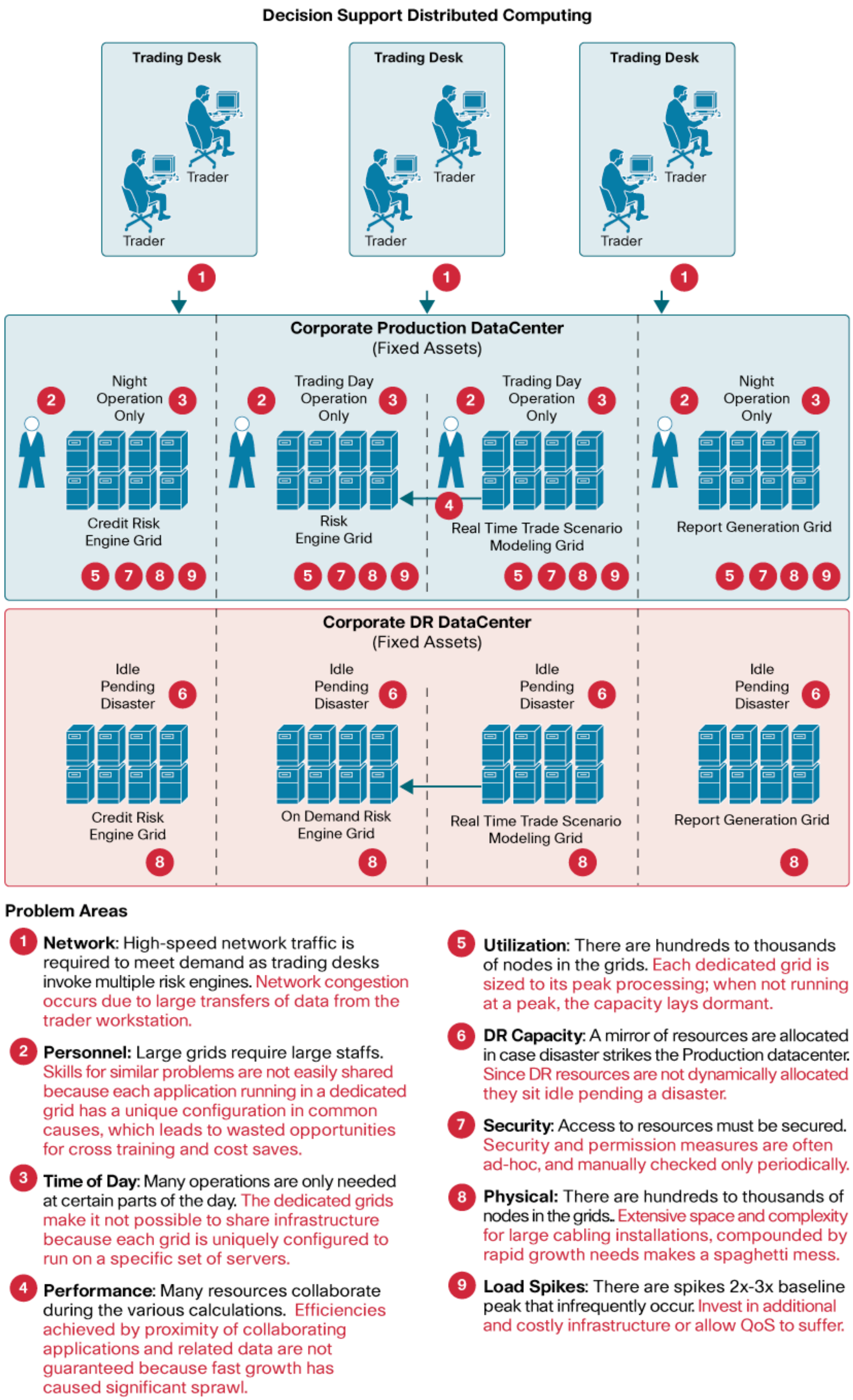
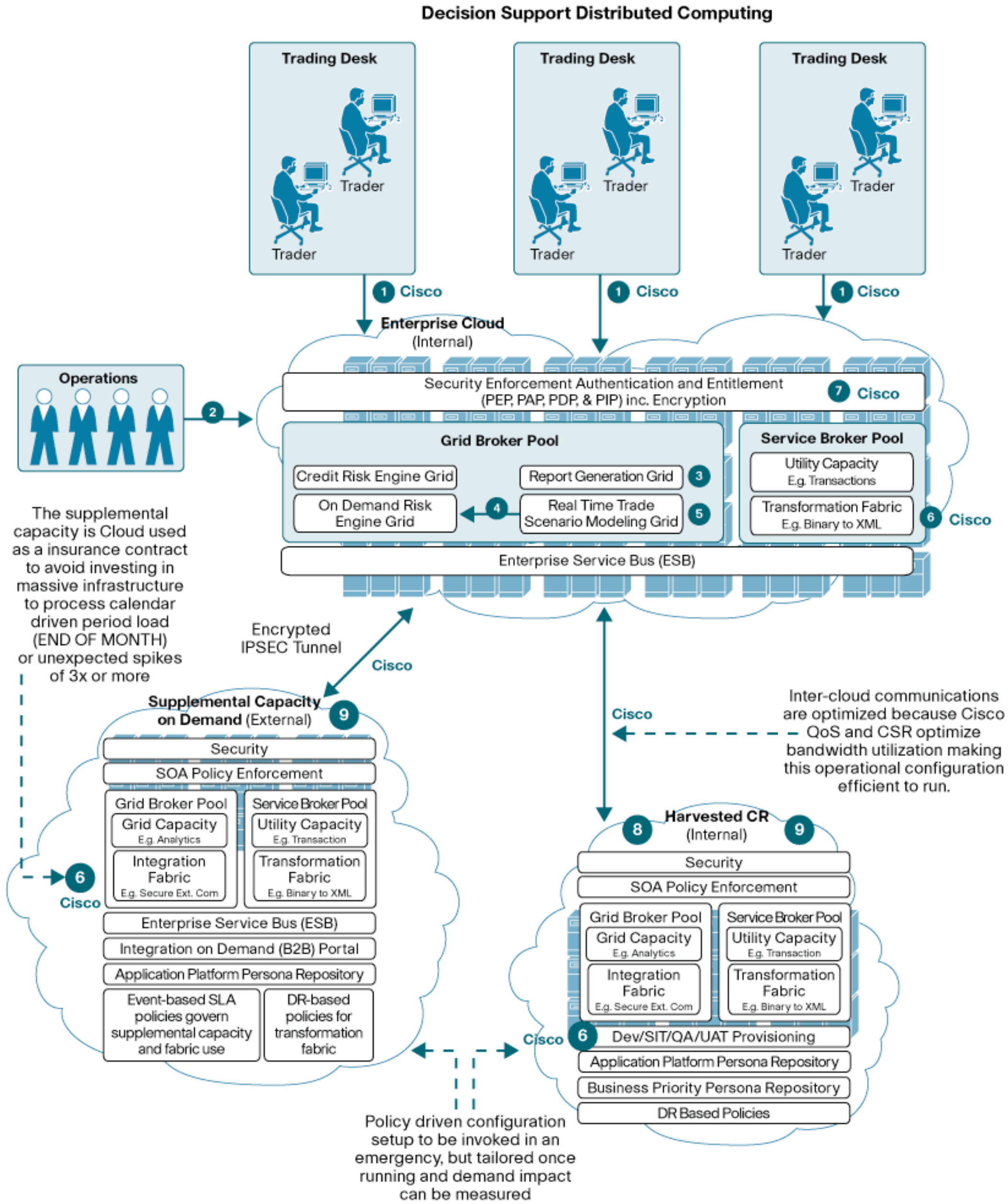


Figure 6. Cloud solution for a Decision Support Grid Computing Utility



Solutions

1. **Network**: High-speed network traffic is required to meet demand as trading desks invoke multiple risk engines. Network utilization is reduced as data and services move into the datacenter.
2. **Personnel**: Large grids require large staffs. Skill for similar problems are easily shared because each application runs on grids dynamically allocated from pools of resources, which improves opportunities to cross train and save money.
3. **Time of Day**: Many operations are only needed at certain parts of the day. Each grid is dynamically allocated and right sized to the demand in real time, which makes it possible to share infrastructure as desired.
4. **Performance**: Many resources collaborate during the various calculations. Efficiencies had by proximity of collaborating applications are possible, as resources are easily localized due to the dynamic operation of the grids and distributed memory (Coherence) / disk (IBRIX) caching to optimize IO.
5. **Utilization**: There are hundreds to thousands of nodes in the grids. Each dynamically allocated grid is sized to its current processing demand; unused capacity is available for other tasks.
6. **Physical**: There are hundreds to thousands of nodes in the grids. The Nexus switches combines network and SAN physical access, reducing complexity, connections, and hardware per server.
7. **Security**: Access to resources must be secured. Cisco Policy Manager and Policy Enforcement Points; ensures WIRE SPEED fine grained auth disallowing invalid requests before they waste resources. Business Policy Driven
   • Policy Enforcement Points (PEPs)
   • Policy Administration Points (PAPs)
   • Policy Decision Points (PDPs)
8. **DR Capacity**: A mirror of resources are allocated in case disaster strikes the Production DataCenter. Since DR resources are dynamically allocated they can be tasked with other work while not needed for a disaster and quickly brought to bear when there is.
9. **Load Spikes:** There are spikes 2x-3x baseline peak that infrequently occur. Dynamically tap supplemental third party or DR cloud capacity to meet spikes.

Application Delivery Use Case

Day in the Life of the Typical Employee in a Retail Brokerage Branch

The retail branch employee depends heavily on the branch workstation, as it supports the user experience for clients while keeping employees in touch with the larger organization. Employees in branches perform many client-facing duties including arranging client advisory meetings, running seminars, and initiating new accounts. Retail branches represent the larger corporation, therefore it is critical that IT supports the branch in these efforts. Keeping employees informed of changes is critical in a larger corporation, thus employee self-service is a vital part of ensuring consistency of process while capping costs for employee benefits.

Requirements to Fulfill the Retail Branch's Needs

Retail branch personnel need to support the client account managers and thus must have some of the same robust applications. In addition they need a suite of applications that the enterprise makes available to take care of procurement, general administrative duties, HR self-service, compliance and reporting, and more.

Problems and Consequences in Meeting that Need

The enterprise needs to provide a large array of applications that while important are not strategic to the firm's revenue-generating goals. The traditional approach is to dedicate staff and resources to these applications in the enterprise data centers. These applications must provide high levels of reliable service and thus become significant investments in infrastructure, software, and support personnel, even though they are not tied directly to revenue. There are problems that persist as a result of this traditional arrangement:

- **Administrative Applications take Significant Resources**: Many applications are not strategic to the business but must run reliably with service performance levels of many client-facing applications. The cost of retooling application platforms to meet growing demand is rarely related to revenue growth. As these applications rarely share base infrastructure, companies must dedicate servers for each one.

- **Unstructured Data Archiving**: Branch desktops and branch servers often serve as local repositories for unstructured data. They can become overwhelmed with this data.
- **Managing Patches and Software Updates**: Desktops with full applications need frequent updates, which takes significant resources and is error prone.
- **Data Center Sprawl**: If archiving for email and unstructured data is pushed to the data centers for all branch activity, then the data center is stressed to manage the significant amount of data, which grows at 40 percent per year in most large enterprises.
- **Collaboration Tools Investments**: Large enterprises will rely on video and web conferencing and other forms of collaboration more heavily as cost pressures in financial services mount. Managing these takes significant data center and infrastructure resources to run correctly. If they are not run correctly they will not be used.

Figure 7. Traditional Non-Strategic Enterprise Applications

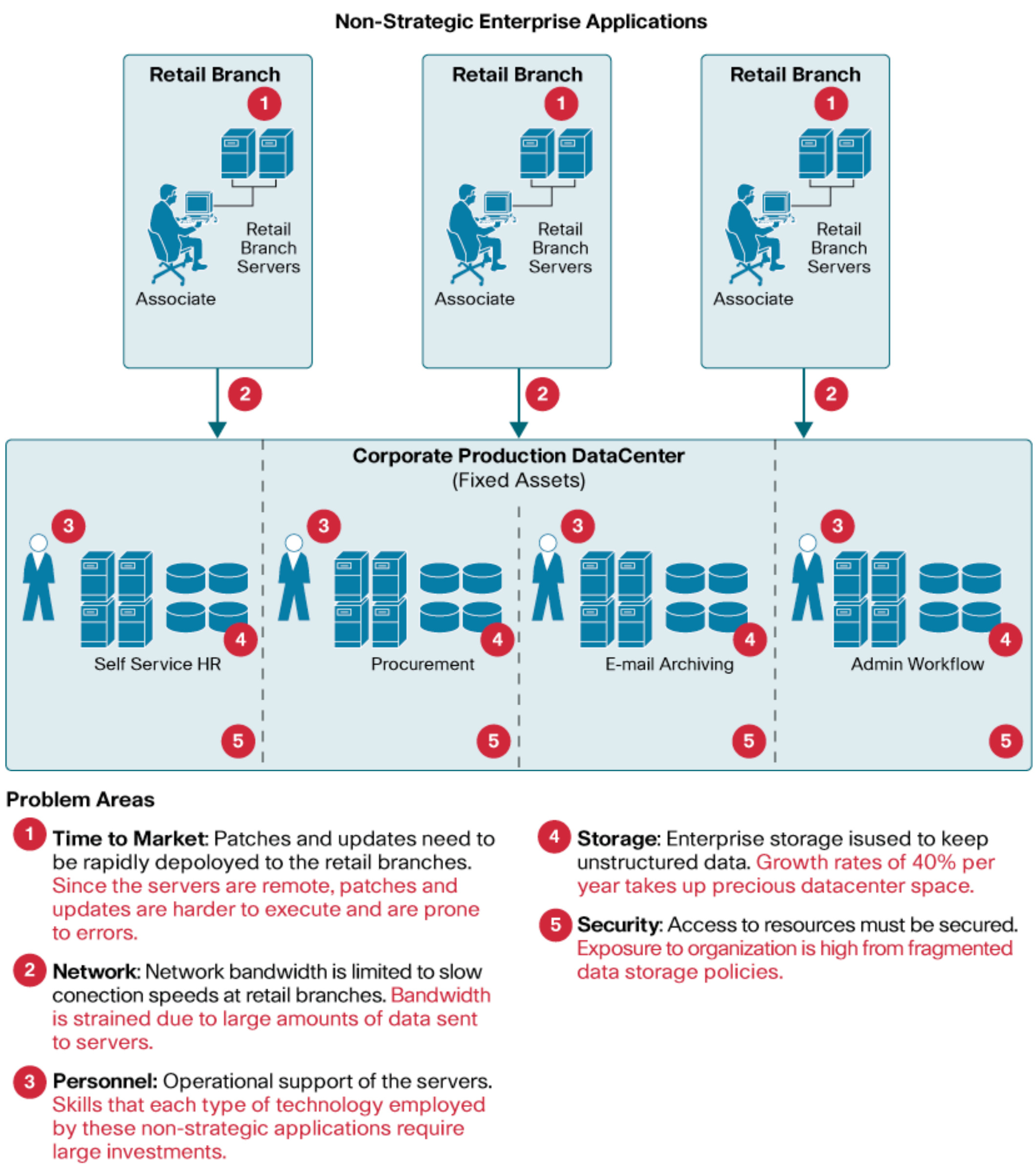
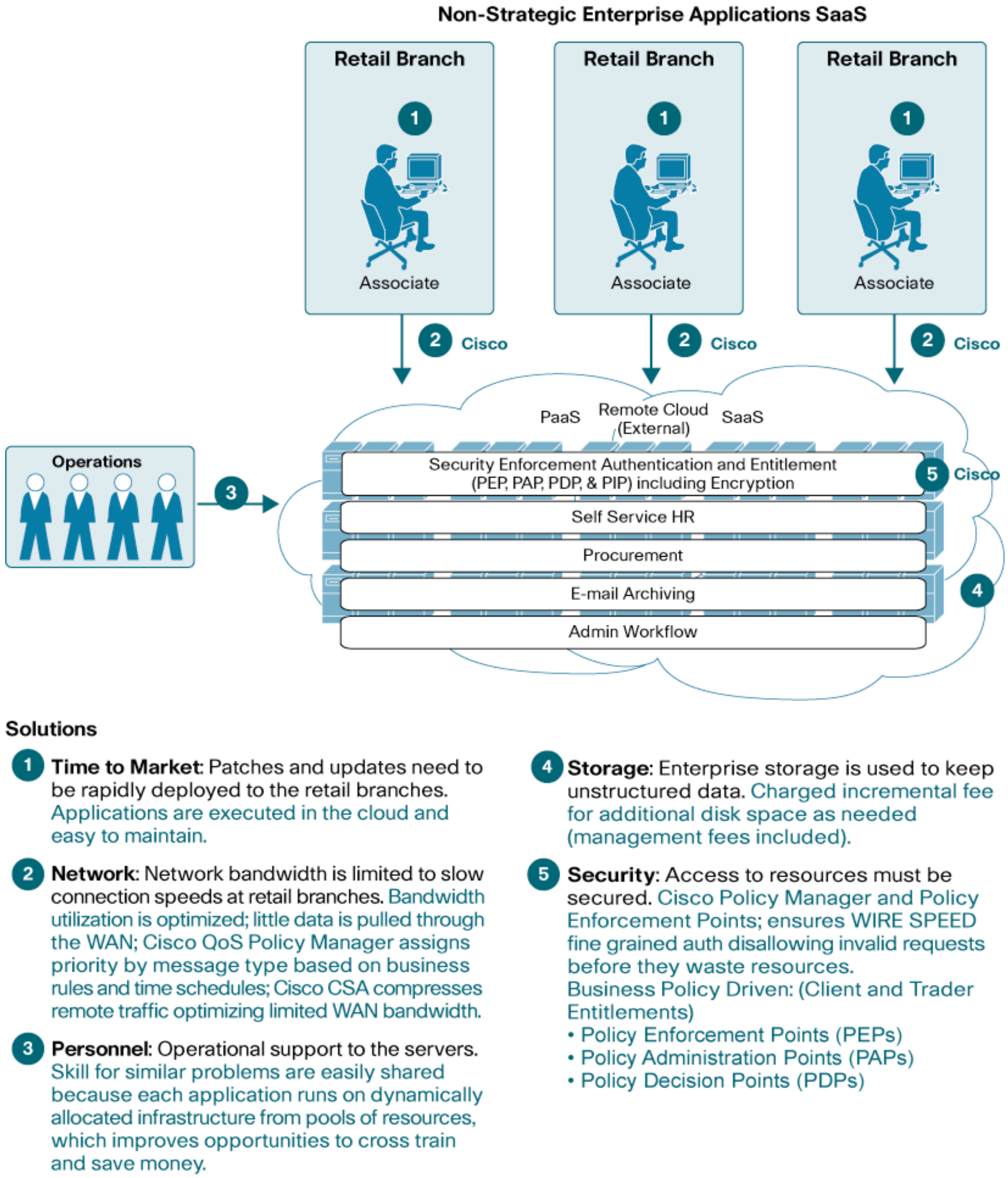


Figure 8. Cloud Solution for Nonstrategic Enterprise Applications SaaS



Work Shift Adjustment Use Case

Day in the Life of the Application Suites in an Enterprise

Many critical applications have operating characteristics that typically lead them to be put on dedicated platforms. During the day they support trading, sales, client account management, and so on. At night, batch processes run that calculate risk across the institution on different levels, and others run settlement and clearing operations. Also required are compliance, audit, asset allocation, inventory, ledger updates, liquidity calculations, and a suite of reporting functions.

Requirements to Fulfill Decision Support System's Needs

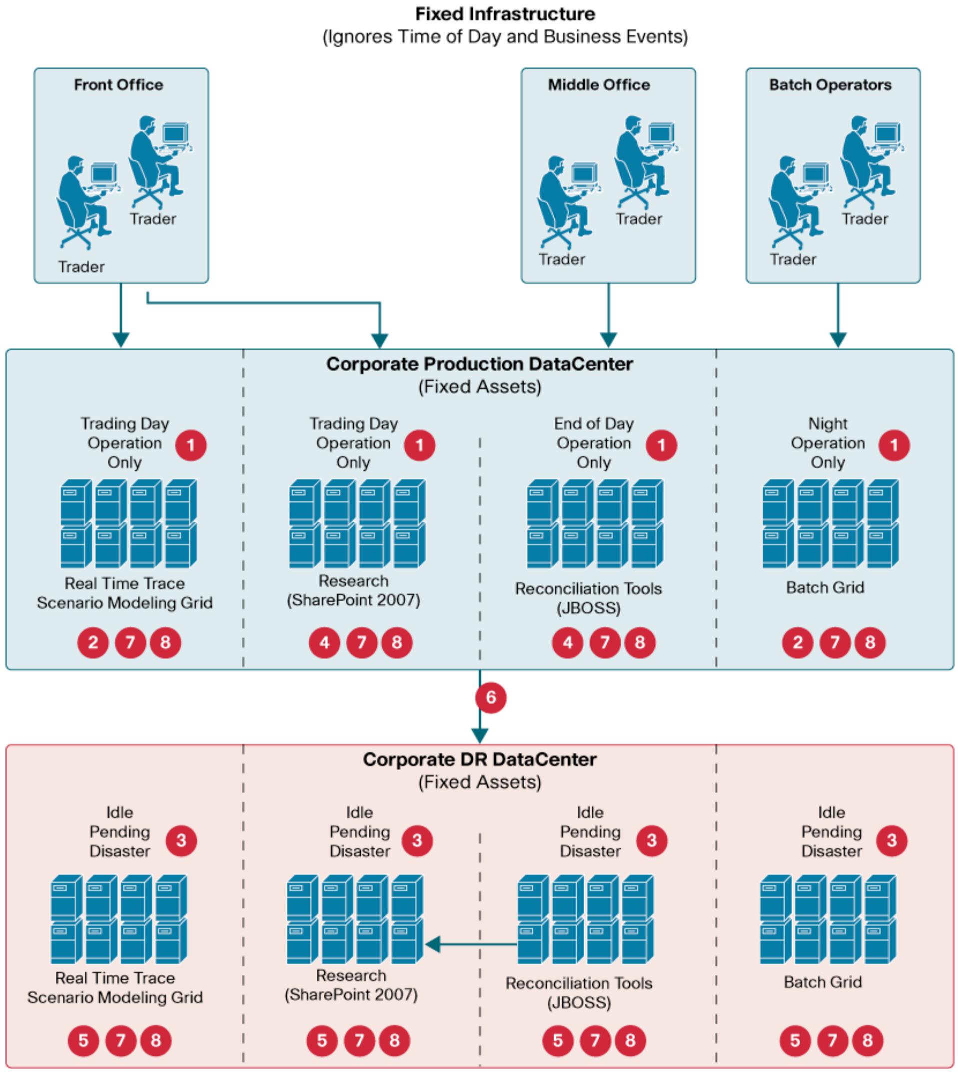
The trading environment requires high-performance operating platforms, but so do functions that support risk, reporting, and other back-office operations. Collaboration tools reduce workflow processing, but at the cost of massive proliferation of unstructured data and use of rich media that requires significant network bandwidth. Daytime requirements of low latency and tight response time are replaced at night by the need to perform massive calculations on gigabyte data sets that require high bandwidth. These calculations must be performed in tight timeframes (2–6 hours) and are continually challenged with increased market volume. The need to finish end-of-day processes in one continent is balanced with the active market on another. Real-time demand outpaces supply, which causes an extreme IT response.

Problems and Consequences in Meeting that Need

The typical response to meeting the different classes of demand for these critical applications has been to build dedicated applications per platform to guarantee stability and performance. But demand continues to increase while nightly timeframes shrink. The result is an ad hoc strategy that overprovisions each platform by two or three times. This creates instability and wastes resources, but none of the dedicated platforms can offload work from the current shift.

- **Data Center Sprawl**: Inefficient use of heavily invested resources is wasteful yet additional resources are always needed to meet application peaks.
- **Unstructured Data Proliferation**: Use of collaboration tools such as Microsoft SharePoint and MOSS increases network demand while creating massive data sprawl.
- **Lack of Business Alignment**: Applications do not grow in proportion to revenue, creating an alignment mismatch that starves agility.
- **Network Utilization**: The proliferation of hardware (CPU, memory, and disk) and the stress of data center floor space cause excessive and inefficient bandwidth utilization.
- **Support Costs**: The skill sets required for such diverse configuration maintenance are not shared across business lines, driving up costs.
- **Idle Capacity**: High-performance resources lie dormant while critical applications are resource starved.
- **Growth Driver**: The regulatory environment drives the need for better risk and reporting forcing individual applications to grow, even as the apparent utilization of resources (when viewed from a 24-hour perspective) continues to dwindle below 25 percent.

Figure 9. Cloud Solution for Nonstrategic Enterprise Applications SaaS



Problem Areas

1 **Time of Day**: Many operations are only needed at certain parts of the day. The rigidity of the software installs makes it not possible to share infrastructure because each grid is uniquely configured to run on a specific set of servers.

2 **Utilization**: There are hundreds to thousands of nodes in the grids. Each dedicated grid is sized to its peak processing; when not running at a peak, the capacity lays dormant.

3 **DR Capacity**: A mirror of resources are allocated in case disaster strikes the Production datacenter. Since DR resources are not dynamically allocated they sit idle pending a disaster.

4 **Specialization**: A heterogeneous environment of application servers is utilized in the environment. Resources are fixed, such that changing the application server type requires a rebuild.

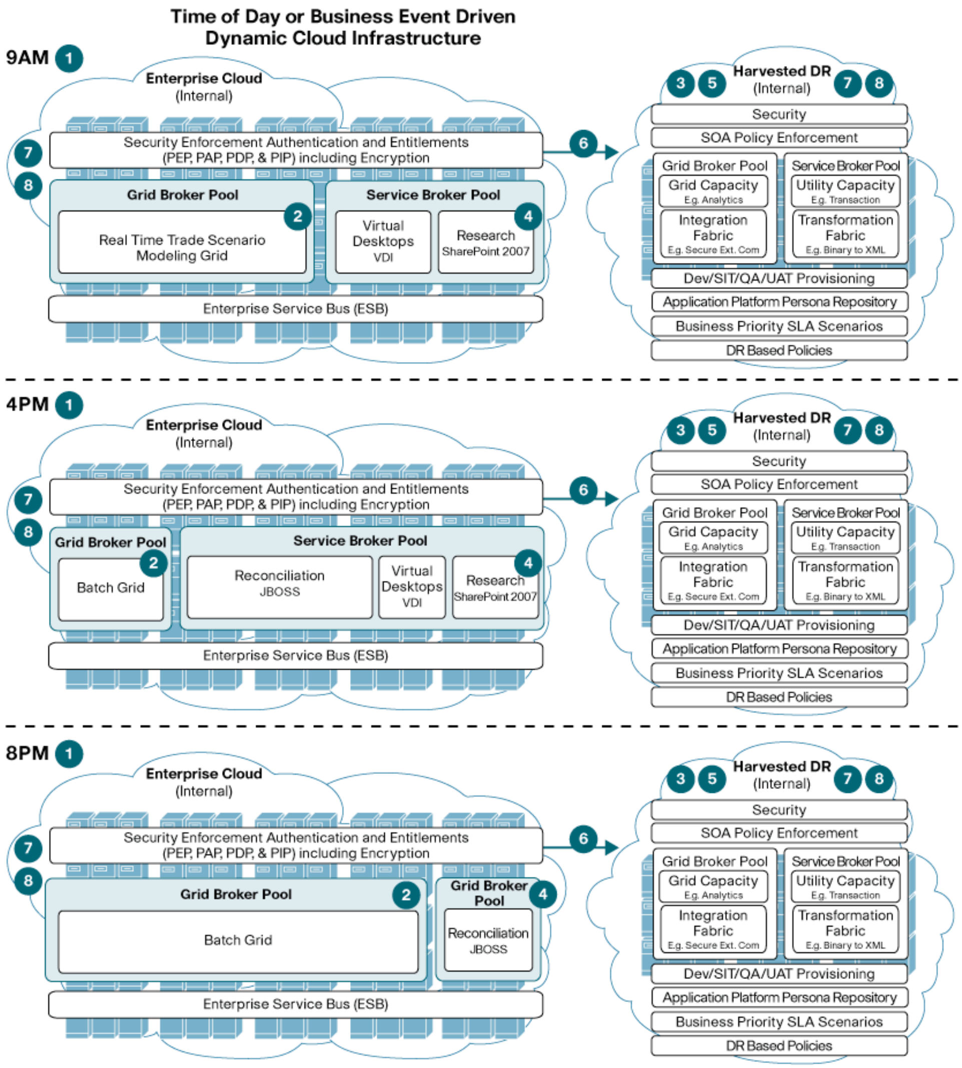
5 **DR Assurance**: The DR environment mut be tested regularly to validate its readiness. Patches and maintenance changes cause differences between DR and Production which are often missed due to the manual time consuming execution of DR testing.

6 **DR Synchronization**: The timely movement of data files and configurations is critical to DR success. Achieving DR synchronization is difficult due to uncoordinated and arbitrarily large data transfers, which takes time to transfer.

7 **Connectivity**: Each server resource needs connectivity to the network. Layer 2 connections are treated independently leaving each resouce to separately manage resiliency making recovery complex. Connections are not virtual machine aware causing contention.

8 **Network Management**: There are 10s to 100s of network switches that need configurations. Each switch must be independently managed which increases complexity and reduces resiliency. Limitd support for Virtual Access Layer and Security Policies Management.

Figure 10. Cloud Solution for Nonstrategic Enterprise Applications SaaS



Solutions

1. **Time of Day**: Many operations are only needed at certain parts of the day. Each grid is dynamically allocated and right sized to the demand in real time, which makes it possible to share infrastructure as desired.

2. **Utilization**: There are hundreds to thousands of nodes in the grids. Each dynamically allocated grid is sized to its current processing demand; unused capacity is available for other tasks.

3. **DR Capacity**: A mirror of resources are allocated in case disaster strikes the Production Datacenter. Since DR resources are dynamically allocated they can be tasked with other work while not needed for a disaster and quickly brought to bear when there is.

4. **Specialization**: A heterogeneous environment of application servers is utilized in the environment. Resources are dynamically allocate and can be spun up as any platform type.

5. **DR Assurance**: The DR environment must be tested regularly to validate its readiness. The automated provisioning and configuration of the DR environment facilitates this process; it is possible to automate this validation.

6. **DR Synchronization**: The timely movement of data file and configuration is critical to DR success. Inter-cloud communications are optimized because Cisco QoS and CSR optimize bandwidth utilization.

7. **Connectivity**: Each server resource needs connectivity to the network. Cisco Nexus switches enable. Layer 2 connectivity across Data Centers by leveraging the Nexus' Virtual Port Channel (vPC) feature. This increases resiliency with less complexity while providing enhanced support for VDI, virtual machines and DR mobility.

8. **Network Management**: There are 10s to 100s of network switches that need configurations. With Cisco Nexus switches as row and rack ends, an entire row can be managed from a single locations, including the Virtual Access Layer and Security Policies thereby reducing complexity and increasing reliability.

Disaster Recovery Infrastructure Use Case

Day in the Life of Disaster Recovery Platforms

Disaster recovery (DR) platforms generally prioritize applications into tiers. In case of disaster the top-tier applications move first. Many of these collaborate heavily, so moving a top-tier application may require lower priority applications to move as well. Tiers are useful in determining what to duplicate in the DR site, as providing 100 percent redundancy for all applications is neither cost-effective nor feasible. Traditionally DR platforms remain idle ready for a disaster event to happen. The idea is that having warm infrastructure standing by minimizes confusion and recovery time. However, demand on the production side continues to grow, and so do software changes. Often it is very hard to maintain a full copy of the current configuration, guaranteeing that confusion will occur despite having the hardware ready.

Requirements to Fulfill Disaster Recovery System's Needs

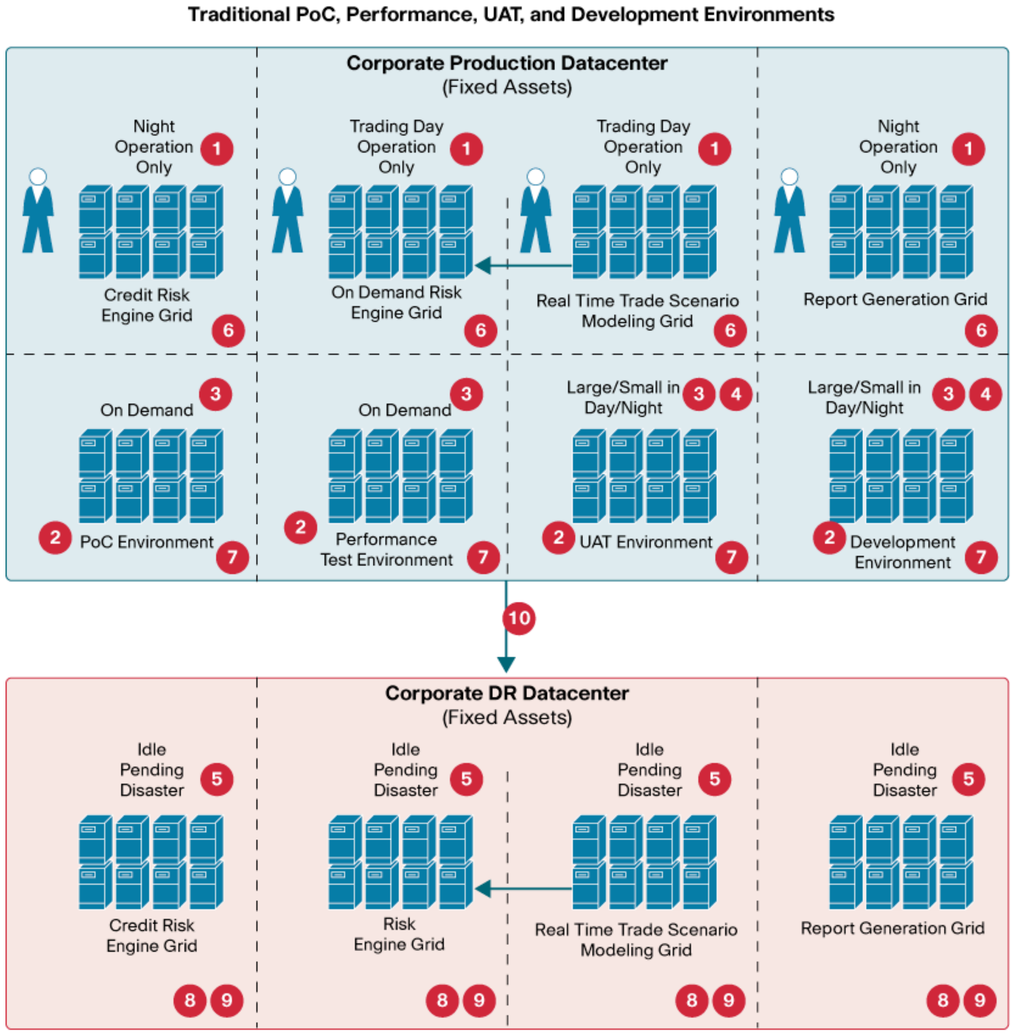
Each application dictates its operational needs, but some needs transcend all applications. Every DR system needs updated start-of-day information delivered every night. Every overnight preparation needs to be run in case a disaster strikes. If the primary application is patched or upgraded or changes configuration, the DR system must receive it. Many changes can occur in an average week, but all changes for all applications must be processed in preparation for the next day. New configurations should be tested at the DR site before the DR needs to be invoked. Finally, the DR site should be tested or even become primary periodically to ensure its readiness. Consideration should be given to using internal DR sites for high-priority applications and external DR facilities for other applications. A mixed DR scenario is possible where multiple sites assume part of the DR responsibility, but that requires DR to be available from multiple locations, which is more complex in traditional operating environments.

Problems and Consequences in Meeting that Need

DR systems have all the normal operational constraints faced by the primary in traditional IT arrangements. What makes DR unique is the extra burdens imposed by the following:

- **Limited Capacity**: Most DR systems have less capacity than the primary; therefore, the performance limits are not well known. Meanwhile, all DR systems are built as dedicated stand-alone systems with no ability to share the infrastructure.
- **Getting End-of-Day Data**: DR sites are remote from the primary but need all changes to files and all data to run at the start of day. Sending files to DR is typically ad-hoc and lacks priority or structure. The amount of data sent on a given night can tax WANs, and there is no guarantee high-priority applications will properly complete their transmission.
- **Poor Utilization**: There is reluctance to make use of remote DR for testing, UAT, or other independent tasks given the risk of not responding to a DR situation.
- **DR Assurance and Environment Matching**: There is no guarantee that all changes from the primary have been properly applied to the DR platform. Testing is not done often enough and periodic swapping of primary / DR sites is rarely considered feasible.

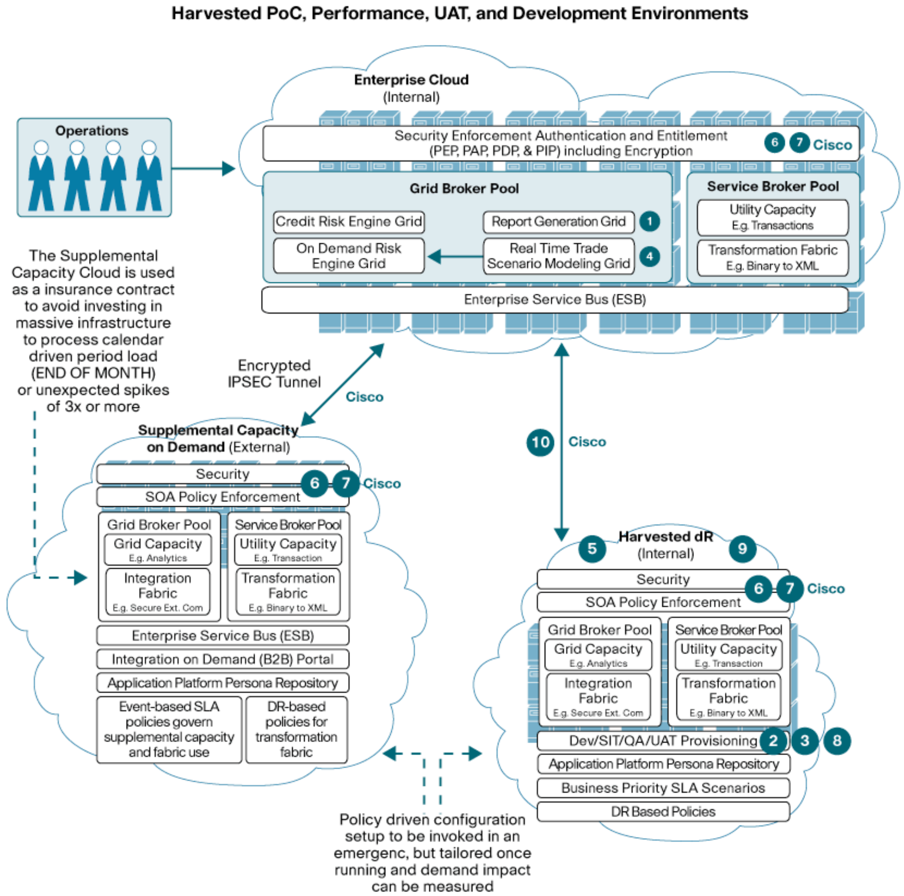
Figure 11. Traditional PoC, Performance, UAT, and Development Environments



Problem Areas

1. **Time of Day**: Many operations are only needed at certain parts of the day. The dedicated grids make it not possible to share infrastructure because each grid is uniquely configured to run on a specific set of servers.
2. **Resiliency**: The availability of the PoC, performance testing, UAT, and development environments is critical to meeting promised development deadlines. A failure of any piece of equipment or misconfiguration can cause large delays while the problem is repaired.
3. **On Demand**: PoC and performance test environments are temporary in nature and are only needed for short periods of time, but must be able to be brought online when needed. Environments are static and costly/time consuming to stand up and tear down.
4. **Utilization**: UAT development environments tend to be heavily utilized in the day, but not at night. Environments are static and can't grow or shrink easily as needs change.
5. **DR Capacity**: A mirror of resources are allocated in case disaster strikes the Production datacenter. Since DR resources are not dynamically allocated they sit idle pending a disaster.
6. **Security**: Access to resources must be secured. Security and permission measures are often ad-hoc, and manually checked only periodically.
7. **Security**: The development teams should have access to a standardized security model to develop against. No security standardization exists so developers create ad-hoc solutions.
8. **Environment Matching:** The test and load environment should replicate production as closely as possible. Due to the cost, this rarely is possible.
9. **DR Assurance**: The DR environment must be tested regularly to validate its readiness. Patches and maintenance changes cause differences between DR and Production which are often missed due to the manual and time consuming execution of DR testing.
10. **DR Synchronization**: The timely movement of data files and configurations is critical to DR success. Achieving DR synchronization is difficult due to uncoordinated and arbitrarily large data transfers, which takes time to transfer.

Figure 12. Cloud Harvested PoC, Performance, UAT, and Development Environments



Problem Areas

1. **Time of Day**: Many operations are only needed at certain parts of the day. Each grid is dynamically allocated and right sized to the demand in real time, which makes it possible to share infrastructure as desired.

2. **Resiliency**: The availability of the PoC, performance testing, UAT, and development environments is critical to meeting promised development deadlines. These environments now run within the DR harvested cloud, which self heals if a failure occurs; Environment is available except for when a full fledged disaster occurs at which point it is not needed.

3. **On Demand**: PoC and performance test environments are temporary in nature and are only needed for short periods of time, but must be able to be brought online when needed. Environments are dynamic and cost efficient.

4. **Utilization**: There are hundreds to thousands of nodes in the grids. Each dynamically allocated grid is sized to its current processing demand; unused capacity is available for other tasks.

5. **DR Capacity**: A mirror of resources are allocated in case disaster strikes the Production datacenter. Since DR resources are dynamically allocated they can be tasked with other work while not needed for a disaster and quickly brought to bear when there is.

6. **Security**: Access to resources must be secured. Cisco Policy Manager and Policy Enforcement Points; ensures WIRE SPEED fine grained auth disallowing invalid requests before they waste resources. Business Policy Driven:
   - Policy Enforcement Points (PEPs)
   - Policy Administration Points (PAPs)
   - Policy Decision Points (PDPs)

7. **Security**: The development teams should have access to a standardized security model to develop against. Cisco Policy Manager and PEPs are a standardized security model.

8. **Environment Matching:** The test and load environment should replicate production as closely as possible. Harvested DR resources that facilitate the duplication of the production environment makes it easy to accomplish this.

9. **DR Assurance**: The DR environment must be tested regularly to validate its readiness. The automated provisioning and configuration of the DR environment facilitates this process; it is possible to automate this validation.

10. **DR Synchronization**: The timely movement of data files and configurations is critical to DR success. Inter-cloud communications are optimized because Cisco QoS and CSR optimize bandwidth utilization.

Figure 13. Reduced DR Footprint with Multiple Production Cloud Locations

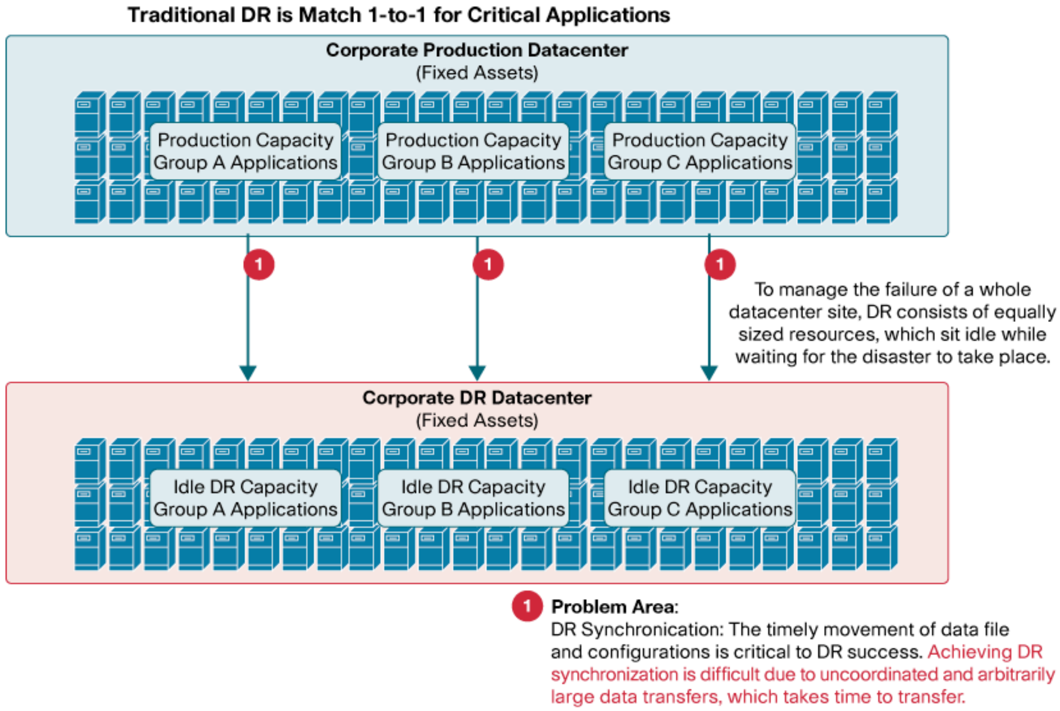
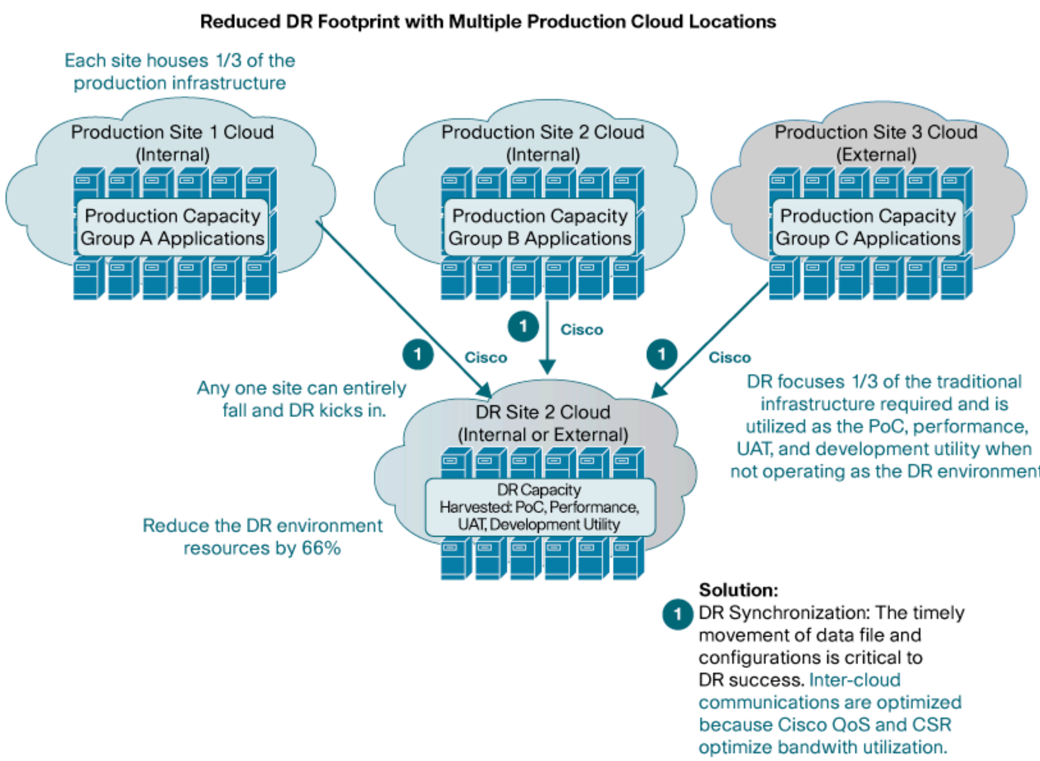


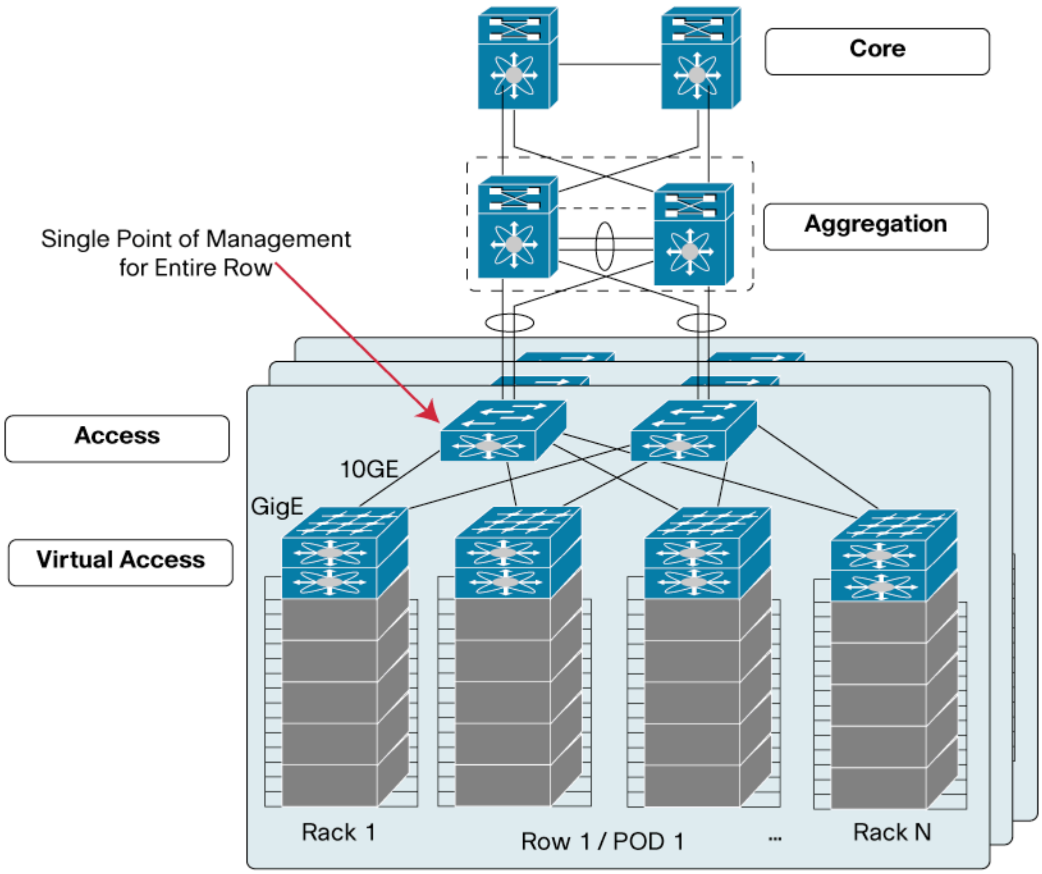
Figure 14. Reduced DR Footprint with Multiple Production Cloud Locations



Design Tips

As cloud environments scale, data center architectures become very important in providing a scalable, reliable, SLA-drive infrastructure. A POD (Point of Delivery) is a module or group of network, compute, storage, and application components that work together to deliver a network service. The POD is a repeatable pattern, and its components maximize the modularity, scalability, and manageability of data centers.

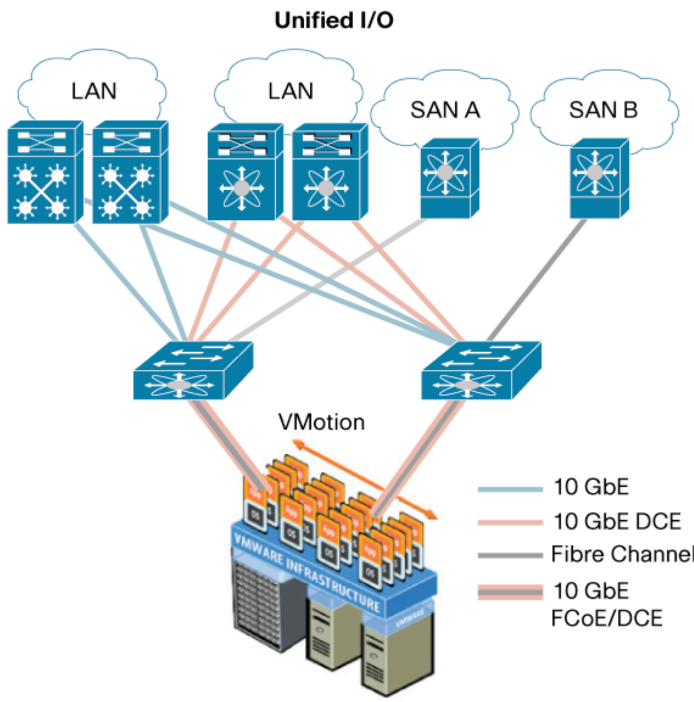
Figure 15. Cloud Point of Delivery Architecture



The POD architecture with the Nexus 2000 Fabric Extender and Nexus 5000 switches solution allows a simple way to reduce complexity and manage an entire row of compute nodes via two central switches. The architecture also provides deterministic latency between any nodes in the entire row vs the tradition rack unit. The Virtual Access layer model allows the Edge of the network to be virtualized to create an environment where independent racks can be deployed on-demand with “zero-touch” configuration at the rack-level. In a cloud environment, scaling workload becomes much easier and OPEX is dramatically decreased through a flexible architecture. The different rows can be consolidated into Nexus 7000 Aggregation switches provide high density compute farms with high availability. The Aggregation switches can be configured with Virtual Device Contexts to logically partition rows where each row is dedicated to a different client. Virtualization of the Nexus 7000 allows complete robust isolation of traffic so multiple tiered services can be offered with no compromise to performance and security.

For cloud environments where virtualized hosts are being used to provide services, one can use the design in Figure 16, which leverages Cisco’s Unified I/O technology.

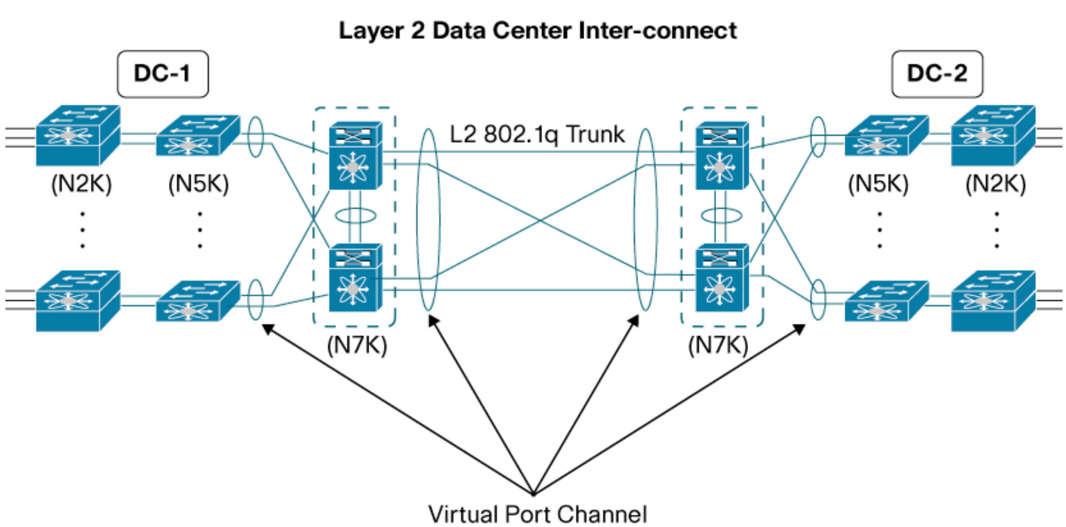
Figure 16. Unified I/O at the Server Level



A Unified I/O link to the host can present multiprotocol traffic to a unified fabric on a single cable. A unified fabric is a single, multipurpose Ethernet transport that can transmit IP and Fibre Channel traffic simultaneously across the same interface and the same switch fabric preserving differentiated classes of service. There are dramatic savings in power, cooling and cabling facilities costs due to less adapter and cables being provisioned to provide connectivity to the various networks in the environment. All the hosts in a particular POD can be equipped with Converged Network Adapters thus provide the capability to provide any storage to any host regardless of physical location. Ubiquitous storage capabilities in the cloud environment are beneficial as demands various storage based networks are required to be addressed.

Another important design tip is the Layer 2 Data Center Inter-connect, especially in the context of Disaster Recovery.

Figure 17. Layer 2 Data Center Inter-connect



In order to provide resiliency for Virtual Machine mobility and clustering services across multiple data centers, there are increasing requirements in the cloud to allow stretched Layer 2 connectivity between sites. In lieu of implementing MPLS and managing the complexity of the protocol and configurations, the Virtual Port Channel (vPC)feature of the Nexus 7000 platform can be used to provide the Layer 2 data center interconnect with which does not depend on spanning-tree for forwarding traffic. Each pair of Nexus 7000 in a vPC domain appears as 1 logical entity from a peering perspective. The vPC feature increases higher resiliency from the server Access layer in the data center since one Nexus 5000 switch can create a logical port-channel across two Aggregation switches. Higher resiliency and quick failover mechanisms enable flexible and resilient architecture to support environments such as VDI and cluster services.

Cisco Cloud Enablement Strategy

Cisco, in conjunction with a few select partners, has pulled together the building blocks needed to rapidly deploy or transform existing resources into cloud-based infrastructures. Delivered as tailored ensembles, they are designed to enable, enforce, and secure the full spectrum of operational qualities that will meet the required service level agreements of a real-time enterprise.

Workload Mobility

- **User Experience**: Tailored entitlement and collaborative tools (instant messaging, video, and so on) to provide rich and fast interactive interfaces that are safe, open architecture, simple, virtualized, and converged.
- **Deployment**: Dynamic real-time and on-demand delivery of resources so they are put to work where needed, as needed.
- **Virtualization**: A broad view of virtualization provides a unified computing platform to include the full breadth of servers, network, and data to dynamically manage abstracted infrastructure, platform and software services, including the use of Scalent VOE, FabricServer, GridServer, VMWare, and other technologies.

Security and Control

- **Security and Integration**: Authentication, role authorization, and fine-grained authorization through Cisco® Application Control Engine and Cisco Policy Manager (PEP/PAP/PDP/PIP) plus ESB functionality to ease integration.
- **Service Execution**: Cisco is a certified partner with DataSynapse, the premier vendor for fabric and grid solutions. This provides a unified fabric for service execution enabling resources to be dynamically repurposed.

Integration and Interoperability

- **Data/Content Delivery**: Comprehensive data center networking (switches, routers, handlers, protocol accelerators, edge computing, and content caching devices) operate at wire speeds to provide flexible high-performance dynamic data delivery within both intercloud and enterprise-class cloud infrastructures to meet business demand.
- **Transaction Processing**: Low-latency, high-throughput, high-availability connectivity with reduced hops and as-needed transformation acceleration facilitates efficient and high-performance intercommunications.
- **Cost to Execute**: Virtualization, abstraction, and simplifying technology (such as Fibre Channel over Ethernet) help to deploy less infrastructure and execute more transactions with fewer people.

Next Step: Contact Cisco

Cisco's approach helps firms quickly establish a strong foundational cloud core. This core can be utilized to produce mass-customizable, reusable building blocks of new IT capabilities that differentiate qualities of user experience while achieving new levels of lean efficiencies. At Cisco, we institute a sustainable culture of continuous innovation for new business capabilities through IT.

For more information on Cisco's solutions for Financial Markets visit http://www.cisco.com/go/financialmarkets.

For more information about the Cisco data center products, please visit http://www.cisco.com/go/datacenter.

For more information about the Cisco data center services that can help you gain the most return from your data center investment, visit http://www.cisco.com/go/dcservices

About the Authors

Tony Bishop is the CEO of Adaptivity Inc.

Sheppard Narkier is the Chief Scientist at Adaptivity Inc.

Alex Pecoraro is the SVP of Product Management at Adaptivity Inc.

Dave Malik is the Director, Solutions Architecture in the Advanced Services organization at Cisco Systems, Inc.

Mihaela Risca is the Global Solution Manager for Financial Markets in the Marketing organization at Cisco Systems, Inc.

Printed in USA C11-518876-00 02/09