WHITE PAPER

CDW PEOPLE WHO GET IT

THE HIGH-PERFORMANCE WAN

With the right optimization tools and acceleration strategies, an organization can keep network traffic from hitting a logjam.

Executive Summary

Wide area networks (WANs) are critical to the fabric underlying all business-critical applications, whether in organizational data centers or spread out to the cloud. But for large, distributed entities, maintaining network links to each branch office can be costly.

WAN optimization and acceleration is not as simple as boosting the bandwidth in a slow office or dropping a single device into the network. Optimizing WANs for best performance requires bringing together multiple networking and security technologies.

Network and security managers must work hand in hand to incorporate data compression and reduction, application optimization, traffic prioritization and bandwidth management, network visibility, load balancing and dynamic routing to the WAN.

The tricky part of network optimization is handling the interactions among multiple devices, product makers and technologies. Because any change to the network can have wide-ranging effects, an organization must be careful with product selection, testing and design. All of these are more important than ever, given the growth in mobile users and the demands they now place on most organizational networks.

TWEET THIS!

Table of Contents

The Network Now

As enterprise IT continues to roll out mission-critical online applications, network managers are becoming key players in ensuring the success of new initiatives. With knowledge workers at branch offices, home offices and roaming the world, the network has become a salient link between users and applications. At the same time, network costs have come under the budget microscope — with pressure to keep monthly WAN spending within limits.

Other networking trends also are driving the demand for a well-managed, high-performance WAN. Data center consolidation and removal of servers from branch offices both bring dramatic savings in time and complexity, yet also increase dependence on the network. Without a server nearby, the network is all that keeps branch offices and remote workers up and running productively.

Finally, applications themselves demand increasing use of the network. Old-school, "green screen" applications — which were stingy on bandwidth and could operate well over low-speed, high-latency circuits — are giving way to AJAX-based web apps, multimedia apps (voice, video and collaboration tools) and apps that integrate large data objects (such as email attachments).

Although organizations can often afford to increase bandwidth as worldwide networking costs drop, latency and reliability problems are not as easily overcome. Simply increasing bandwidth without also reducing latency does not provide significant benefits to web-based apps.

The consumerization of networking, driven by huge requirements for Internet access in homes, has shifted data communications away from high-cost, high-reliability dedicated circuits to low-cost, moderate-reliability services based on Ethernet, cable modems, digital subscriber line (DSL) services and wireless.

Prices for bandwidth have dropped 100- to 1,000-fold over the past 20 years, but the result has been lower-quality bandwidth — more jitter and unpredictable latency; greater contention during peak periods; and a lower level of service, driven by the lower costs demanded by consumers.

These trends in network growth and application profiles have driven many network managers to react by updating enterprise networks — increasing bandwidth, taking advantage of lower-cost services and expanding visibility. But many standard upgrade paths have undesirable consequences — the "more of the same" approach may take the network two steps forward but also one step back.

Building out enterprise WAN service requires a new approach. Growth requires an additional layer of visibility, control and optimization that can take the network forward to meet the needs of newer, bandwidth-heavy apps.

Common approaches taken by network managers and potential consequences:

Trend	Action	Possible Consequences
Greater use of bandwidth by a larger number of applications and larger data object sizes	Increase bandwidth of circuits	Application contention by a greater number of applications; disappointing results if latency is not also reduced
Lower costs for many WAN circuits	Use lower-cost services	Higher level of downtime; uncontrolled jitter and latency
Heavier use of multimedia applications, including voice and video, online collaboration and real-time communication	Divide networks into real-time and non-real-time segments, if possible. Do this through service-level agreements (SLAs) for various traffic types.	Underutilized circuits; increased network complexity; heavier management burden
Overall heavier use of network by all segments of the workforce	Closer attention to network monitoring	Disconnect between what policy allows and how network is used

Fortunately, a wide variety of WAN optimization and acceleration tools, protocols and products are available to network managers to increase reliability, provide visibility, control usage and squeeze every bit of performance out of the enterprise WAN. Network managers can use these

IP Traffic Continues to Grow

A significant driver of the need for WAN optimization and acceleration tools and technologies is pure traffic: more bits across the network — all the time. The *Cisco Visual Networking Index: Forecast and Methodology, 2011–2016*, attributes acceleration in network use to the synergy of multiple fundamental enablers:

- More connections to the global Internet, heavily pushed by mobile networking and smartphones
- More applications, including long-form video streaming, Voice over IP (VoIP) and video conferencing, traditional collaboration and file-sharing tools, and consumer gaming
- More time spent running applications and using the network
- Higher bandwidth available to end users, including business, consumer and mobile networks

These factors work together to increase and accelerate the use of WANs.

Cisco estimates that global IP traffic will grow at a compound annual rate of 29 percent from 2011 to 2016, and that peak traffic load will jump fivefold by 2016.

tools, individually or together, to meet the needs of their organizations' users and apps.

7 Core WAN Solution Sets

Proper WAN optimization and acceleration requires bringing together the right tools to solve the problems of network growth in a holistic way. Network decision-makers will quickly discover that no single box or product solves all issues. Each network environment will have different requirements that lead to the integration of multivendor products to provide a complete solution.

The chart below identifies some of the problems associated with large WANs and the technologies typically used to solve them. In the following sections, this white paper will highlight common networking challenges and provide details about the seven chief solution sets to remediate them.

Problem Area	Technology Solutions
1) Insufficient bandwidth	Data compression and data reduction, including caching and deduplication
2) Misbehaving WAN applications	Application optimization, web-layer controls (web application firewalls)
3) Application contention	Traffic prioritization, quality of service (QoS), bandwidth management
4) Network visibility	NetFlow, IPFIX, sFlow, jFlow, cflowd, rFlow, AppFlow, proprietary tools
5) Network reliability	Dynamic routing, network link balancing
6) Application reliability	Local load balancing, global (multiple data center) load balancing
7) Network abuse and misuse	Access control tools (next-generation firewalls), antimalware, intrusion prevention, web security gateways and proxy servers

Challenge 1:

How can reduction and compression technologies solve insufficient bandwidth?

The proliferation of apps assuming LAN speeds has put pressure on WAN sites that don't enjoy unlimited bandwidth.

Data compression is a key technology for reducing such stress. It works well for most data, except for real-time multimedia – such as Voice over IP (VoIP) or video conferencing – which are already compressed and can't benefit from simple compression techniques.

WAN optimization products implement compression in many ways, including:

- **Standard compression:** This method takes streams of data and sends a reduced version of the content across the circuit, saving bandwidth. Standard compression in a WAN environment has many subtleties, including the choice of algorithm, how compression works across streams, and the interaction between compression and encrypted traffic.
- **Caching:** This technique reduces data by maintaining a stored version of recently requested data objects (typically, files or email attachments) at the remote side of the connection. If a data object is requested a second (or third, or fourth) time and it is in the cache, then that copy is returned, eliminating the need to re-transmit the object from the central site to the remote site.

  Caching is especially useful in environments where file sharing is done across the WAN (using programs such as SharePoint, protocols such as Server Message Block), or where the email server (typically Exchange) is located at the central site and not the remote site.
- **Deduplication:** This approach reduces data by detecting duplication in streams of bytes. Deduplication is a term from the world of storage and backup systems – think of all those nearly identical copies of the C:\WINDOWS\DIRECTORY in a typical backup server and it's easy to understand the benefit of deduplication.

The actual details pertaining to which of these algorithms is used and whether the vendor calls it caching or deduplication is mostly irrelevant. One important difference is that caching nearly always requires a hard disk of some sort to hold cached data, while deduplication is handled on the fly without any persistent storage.

There Is No Compression Without Decryption

All types of applications have been including Secure Sockets Layer (SSL) encryption as a default, even within enterprise networks.

Security teams have been pushing for higher end-to-end security for decades, and application development and system performance are finally catching up. Encryption, however, comes with a downside: There's no good way to optimize, compress or cache information from an encrypted connection.

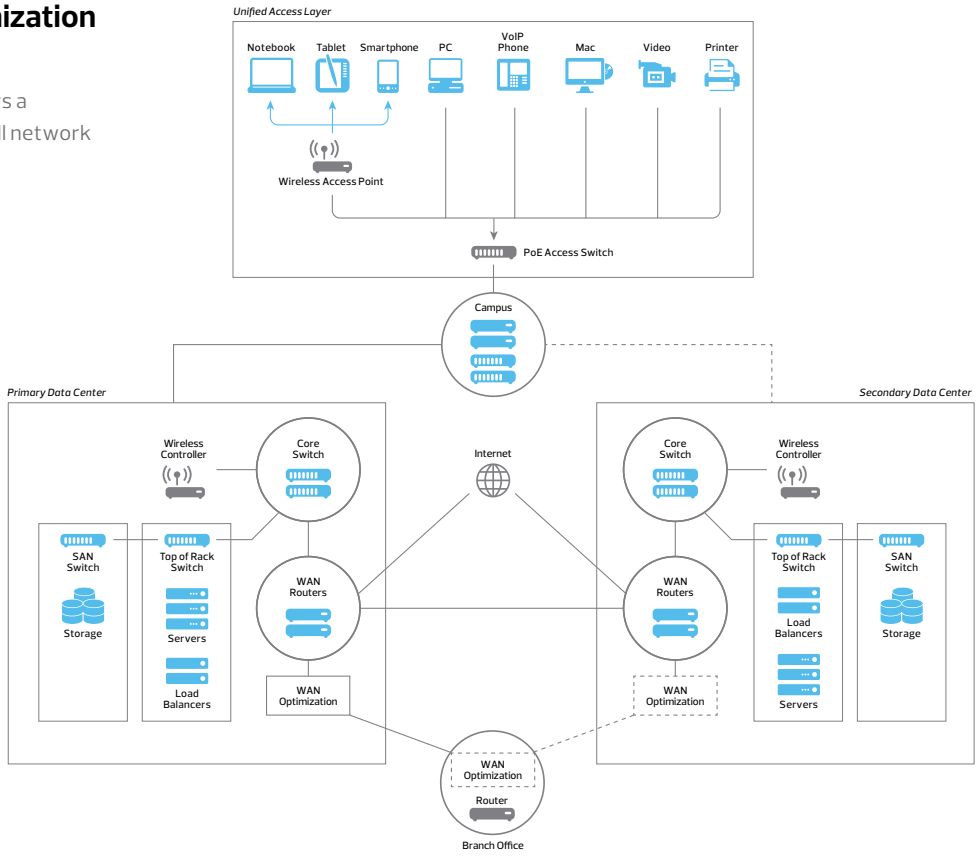
The common use of encryption means that network managers must configure a WAN compression tool to decrypt the SSL traffic – let it do its particular compression magic and then re-encrypt the data – all without slowing traffic down. Without full-speed decryption, WAN compression techniques won't work with today's applications.

In addition to the obvious performance demands from all the encryption, SSL adds some management overhead, as organizations without a solid public-key infrastructure (PKI) in place will need to add that burden to their WAN optimization project.

Standards and cryptographic software and hardware for implementing encryption are widely available. While easier to employ, software-based encryption requires installation onto the computing device and may slow it down. Hardware-based encryption is generally faster than software-based encryption and typically offers more integrity.

Network Optimization Solution

WAN optimization plays a critical role in an overall network optimization solution.



Network managers interested in data compression to reduce bandwidth can evaluate products only by putting them into place in their own networks and comparing the results. The most important detail about data compression is that it requires two devices, one on either end of each WAN circuit or virtual connection (such as over a virtual private network or VPN connection across the Internet).

Compression product makers have tried to mitigate the need for deployment and management of hardware by providing compression devices as virtual machines. The idea is to offer compression software that runs directly on end-user devices, and to introduce many other WAN optimization and acceleration techniques into their products to provide more all-in-one solutions.

Challenge 2:

How can application optimization solve app issues?

Although compression techniques can provide dramatic performance increases, optimizing apps to run over WANs offers benefits far beyond simple compression.

Application optimization can often be provided by the same hardware used for compression, but there is a key difference: Application optimization requires just one device next to the app server. Because the application optimization directly affects web traffic, optimization benefits all app users, not just WAN users.

Examples of application optimization and corresponding benefits:

Application Optimization	Benefit
Smarter use of browser objects such as JavaScript	Many application developers sloppily have the browser re-download JavaScript and other browser objects, such as style sheets, each time a different page is referenced. Application optimizers can rewrite pages on the fly to ensure that these large objects are cached in the client browser. Reordering objects can also make pages render faster, giving a better user experience.
Compression and optimization of content and images	Web browsers internally support compression without requiring any add-on software; most web servers don't bother to compress objects. Simply compressing on the fly speeds access and reduces network load. Images may also be optimized and reduced in size based on the device being used. For example, a smartphone doesn't need the 1024x768-pixel version of each image.
HyperText Transfer Protocol (HTTP) extensions and support for emerging standards such as the SPDY protocol	HTTP, the protocol underlying Hypertext Markup Language, is notoriously inefficient. Acceleration hardware can help to interleave connections and increase the parallelism to speed access over high- latency, low-bandwidth network connections. SSL offload to the optimization-acceleration device can also speed loaded app servers.

Traditionally, application optimization was the realm of a family of products called application delivery controllers or ADC (formerly known as load balancers). But network product makers have migrated these techniques into other devices as well.

Challenge 3:

How can traffic prioritization and bandwidth management reduce application contention?

Voice and video multimedia apps require fairly constant and predictable bandwidth among simultaneous users — if the number of users is known and controllable, which it often is not. Other apps, such as email and web-based programs, tend to be more bursty in their bandwidth requirements.

A cross-section of techniques to provide sophisticated bidirectional traffic management:

Management Technique	Result
Transmission Control Protocol modification on the fly	By changing TCP window sizes and delaying TCP acknowledgments, individual apps can be more tightly managed and controlled.
Application intelligence for User Datagram Protocol apps	UDP-based apps, such as voice and video, can't be easily flow-controlled the way TCP apps can. By knowing more about the internals of a UDP app, WAN optimization devices can perform call admission control — that is, block calls that would overdrive available bandwidth — and even change the requested coder–decoder and sampling frequency for a voice or video call, reducing total bandwidth consumption.
Subdividing apps	Some apps mix both delay-sensitive and bulk traffic over the same connection. WAN optimization devices may be able to break out multiple types of traffic and give different priorities to each type based on deep knowledge of the internal functions of the app.
App identification	Differentiating between business and recreational apps (such as collaborating via SharePoint versus video streaming from YouTube) goes deeper than looking at port numbers. By directly identifying actual apps, WAN optimization devices can provide granular insight and then limit or guarantee bandwidth as required to meet project objectives.
Time-of-day awareness	Although enterprise data centers run around the clock, many branch offices are 9-to-5 operations. This provides the opportunity to use bandwidth — generally bought on a 24-hour basis — differently during an organization or office's off-hours. Log transfers, backups, software updates and other maintenance activities can all be pushed to off-hours and will benefit from different bandwidth management rules, if the WAN-optimization device supports this. Off peak traffic can also result in cost savings.

In a fully managed hub-and-spoke network, quality of service (QoS) mechanisms can be used to guarantee particular bandwidth and prioritization for each app. But few enterprise networks look like this any more. Multiple data centers, branch-to-branch communication and the use of generally unmanaged circuits (such as Internet, wireless and shared services) have reduced the ability of simple QoS mechanisms to guarantee acceptable app performance.

WAN optimization and acceleration projects also now require some management of bandwidth between sites. Simple mechanisms, such as those found in common edge firewalls with unified threat management (UTM), are not sufficient for the complex requirements of a mix of apps and topology.

Bandwidth management can be particularly trying because true bandwidth management works well only in the outgoing direction for each site. Once the packets have come into a site, they've already consumed bandwidth and pushed out other apps that might have been more important.

Simply dropping packets that exceed predefined limits won't work in most situations. WAN optimization and acceleration vendors have come up with a variety of techniques to provide sophisticated bidirectional traffic management.

Challenge 4:

Can standards-based tools provide network visibility?

Most WAN optimization techniques try to improve service with limited resources by controlling use of those resources. But a significant step toward any WAN optimization and acceleration project depends on gaining network visibility. Therefore, it's imperative that the network management team answer a few questions about apps used within its organization:

- What apps are being used?
- Who is running them and when?
- How much bandwidth do they use — individually and collectively?
- What types of errors are occurring?
- What response times are users experiencing?
- Which systems are the top talkers and which are the top listeners?

The old reporting categories must be modified because visibility in current WAN environments involves far more than merely tracking IP addresses and ports. Network managers with monitoring systems built 10 years ago may need to start over. True network visibility extends up the stack to identifying real people and real apps.

Without strong visibility into the metrics listed above (and more), no WAN optimization and acceleration project can be successful. Control of the unknown simply leads to frustration

and confusion, while good visibility into network and app use can also provide metrics to measure overall project or program success.

Although proprietary systems exist, the Internet Engineering Task Force's (IETF) new Internet Protocol Flow Information Export (IPFIX) standard is the best place to start. Many network devices — including switches, routers, firewalls, WAN optimization controllers (WOCs) and application delivery controllers — will send IPFIX and NetFlow data to a management system. Where no IPFIX data is available, both open-source and commercial hardware and software IPFIX and NetFlow exporters are available to give visibility into unencrypted network traffic.

The benefit of choosing IPFIX and NetFlow is that it represents a standard approach, which means that an organization will be able to gain visibility into different components mixed and matched on its network. There's no need to buy a probe for every office if other devices are providing the same information, for example.

Challenge 5:

Can link balancing and dynamic routing improve network reliability?

Although service-level agreements (SLAs) can set expectations, network managers must prepare for the inevitable link downtime that any WAN will experience. When business-critical apps are used over the network, most organizations choose to use dual links into each of their sites to minimize blockages created by traffic peaks or network problems.

Simply having multiple links doesn't ensure high availability, as some mechanism must be in place to use the links. If VPN tunnels are in place, some organizations use dynamic routing protocols such as the Open Shortest Path First (OSPF) protocol (or, for smaller networks, the Routing Information Protocol or RIP) to make use of dual links.

Having two links on at all times always prompts a return on investment question: How can we use both links and still get the most network for the dollar? WAN optimization and acceleration vendors have introduced a variety of techniques to balance traffic over multiple network links, with varying levels of success. Because TCP/IP networks have their own routing protocols, attempts to force traffic to take a particular route or to signal a route to upstream devices (such as Multi-Protocol Label Switching or MPLS routers) are often complicated and create brittle networks.

While the idea of using as much of two circuits as possible is attractive from budget and philosophic points of view, network managers should very carefully evaluate any vendor proposal to perform outbound load balancing or dynamic link selection. Experiences with this type of load balancing have not been positive for all enterprises. In some cases, these types of technologies have required very specific network configurations for correct operation, and may end up creating more problems than they solve.

Challenge 6:

Can load balancing improve application reliability?

Although app reliability is not necessarily a WAN-specific concern, the importance of enterprise apps emphasizes the need for more sophisticated types of load balancing and high-availability strategies that stretch across data centers.

Traditional load balancing uses a Layer 2 or Layer 3 device as the front-end to a series of systems offering an identical service. As requests come in to the load balancer, it makes a decision based on a predetermined algorithm and passes the request on to whichever system is selected.

(Note, current load balancers, also called application delivery controllers, operate all the way up the protocol stack to layer 7 of the OSI model.)

The load balancer then manages state information so that further requests from the same client are all directed to the same system. The algorithm chosen can be as simple as a round-robin process or it can be more sophisticated, taking into account CPU utilization, response time and other factors.

Originally, the goal of most load balancers was scalability — the ability to handle a greater load than any single server could manage. Over time, the goal has changed. Now, the low cost of server hardware has led many organizations to use load balancers simply for reliability. With two (or more) servers available, uptime can be extended and maintenance windows shortened, even if the load can reside entirely on a single server.

Global load balancing is the industry term for load balancing that takes place across data centers. In this case, a single load balancer cannot be in the path between the end user and the app server. Therefore, some other technique must be used to direct the client to the selected server. As with local load balancers, the approach can be as simple as an active-passive or round-robin process, or it can be more complicated and include techniques such as geolocation to find the closest server to the requesting client.

Although people have been talking about global server load balancing for a decade or more, network managers should be aware of one important fact: Global server load balancing is not a solved problem. Because of the way that the Internet, Domain Name System servers and web browsers work, there is no guaranteed reliable approach to providing high availability across multiple data centers.

Several techniques have been tried, including DNS-based load balancing and Border Gateway Protocol (BGP) load balancing

(using Anycast addresses), but no approach works 100 percent of the time in 100 percent of the possible failure cases. Indeed, the numbers aren't even close to 100 percent. For every global load-balancing technique discussed, there are many potential places where load balancing will not deliver the desired results.

Load balancing is usually provided by dedicated software or hardware, such as a multilayer switch or a DNS server. Many experts consider the distinction between hardware and software load balancers to be no longer meaningful.

Where WAN optimization and acceleration solutions can be found:

Technology	Most Commonly Found In ...	But Also Available In ...
Data compression and reduction	WAN optimization controllers (WOCs); discrete hardware appliances or software-based virtual appliances	Some functionality may be available in web security gateways, but pure data compression and reduction are not often found in other product spaces.
Application optimization	Application delivery controllers (ADCs), load balancers	WOCs often include some application optimization features. Web application firewalls are a separate niche.
Traffic prioritization and bandwidth management	Quality of Service (QoS) and visibility products	WOCs often include traffic prioritization; UTM products and next-generation firewalls generally include basic bandwidth management and prioritization.
Visibility (IPFIX and NetFlow)	Stand-alone IPFIX and NetFlow exporters	High-end WOCs, firewalls and routers all generally will act as IPFIX and NetFlow exporters. In most cases, no stand-alone exporter will be needed.
Routing and link balancing	Branch and edge firewalls or combination router–virtual private network (VPN) devices	Stand-alone edge routers all generally have this capability, but the location of the router outside of the firewall (which prevents it from seeing into encrypted VPN tunnels) pushes this feature into whatever device handles VPNs for the branch.
Load balancing	ADCs	Some firewalls offer this feature as well, although rarely with enterprise-class capabilities.
Security features; use and misuse controls	UTM products and next-generation firewalls	Web security gateways and proxy servers may include limited web-focused features. Stand-alone IPSes are rarely used in the branch when UTM or next-generation firewalls are available.

Challenge 7:

What approach should be used to integrate security tools?

WAN optimization is usually considered a largely technical exercise, the goal of which is to get more value out of each dollar spent for connectivity. Many network managers now take a more holistic view of network use, and therefore look to security-focused products to help them control overall use of both enterprise and Internet apps. Because most WANs already have a firewall device at the border of each remote site, these devices may be called upon to provide more than simple firewall and VPN services.

A typical example might be streaming audio and video controls — either outright blocking of such access (using UTM features such as URL filtering) or limiting bandwidth so that streaming multimedia does not affect other business-critical apps running across the WAN.

Security device manufacturers are bringing many branch management features to their edge devices, including URL and content filtering, app identification and control (so-called next-generation firewall features), bandwidth management, intrusion prevention and antimalware.

Network managers should consider including the capabilities of branch firewall devices in their overall network optimization plan for several reasons.

First, these devices are typically already in use, so activating additional capabilities may be as simple as a few mouse clicks or a low-cost subscription add-on.

Branch firewalls are key parts of the WAN. What's more, changes to traffic profiles or traffic types will also affect the operation and capabilities of the firewalls.

Evaluating Optimization and Acceleration Products

To get your organization started, the seven optimization and acceleration solutions noted are mapped out in the chart at left, detailing which solutions are commonly found in which tools and if those features can be found in alternative technologies. Network managers can use this table to prioritize the technologies they need most in their networks and then work to spec out the minimum number of boxes to solve the maximum number of pressing problems.

Because the WAN toolkit available to network managers is composed of many overlapping products, it's helpful to separate out the seven main functions that are used in building enterprise WANs.

Unfortunately, no vendor has been able to cover all the bases with a single product without huge compromises, despite the "slideware," white papers and procurement check lists. At the same time, there is considerable overlap between different vendors and products.

CDW: A Network Optimization Partner That Gets IT

CDW offers a wide selection of network optimization solutions designed to increase the speed of access to critical applications and information across the enterprise. WAN optimization refers to a comprehensive set of techniques for increasing data transfer efficiencies across wide area networks.

Application delivery solutions enable organizations to simplify management of applications at remote sites by hosting critical apps within their main data center and optimizing the WAN to reduce latency to outlying facilities.

Employing network optimization solutions can help:

- Reduce application latency to remote end-users
- Create multiple pathways to ensure application availability
- Centralize the network environment
- Decrease operating and management costs
- Maximize bandwidth utilization
- Postpone the need to upgrade WAN bandwidth
- Improve disaster recovery position by speeding backup and data replication over the WAN

Your CDW account manager and solution architects are ready to assist with every phase of choosing and leveraging the right WAN optimization solution for your IT environment.

Our approach includes:

- An initial discovery session to understand your goals, requirements and budget
- An assessment review of your existing environment and definition of project requirements
- Detailed manufacturer evaluations, recommendations, future environment design and proof of concept
- Procurement, configuration and deployment of the final solution
- Ongoing product lifecycle support
- Availability of on- and off-premises network managed services

To learn more about CDW's network solutions, contact your CDW account manager, call 800.800.4239 or visit CDW.com/network

riverbed
Think fast.™

Riverbed® WAN optimization solutions liberate organizations from common IT constraints by increasing application performance, enabling consolidation and providing visibility.

CDW.com/riverbed

cisco Gold Partner

Using the Cisco® Application Networking Manager (ANM) to manage Cisco application delivery controller (ADC) infrastructure provides a number of benefits:

- Increases productivity by helping you delegate tasks beyond networking and operations staff
- Decreases risk of service outages due to syntax or processing errors
- Reduces configuration sprawl by helping you standardize device configuration

CDW.com/cisco

BlueCoat

Blue Coat MACH5 goes a step beyond other WAN optimization solutions, which focus exclusively on data applications. With Blue Coat MACH5, you can leverage specialized optimization features for data, video, cloud and web applications — even in IPv4, IPv6 or mixed environments.

CDW.com

TWEET THIS!

CDW® PEOPLE WHO GET IT™