White Paper

4 Steps to Surviving Big Data

As big data projects slam network performance, network managers must be ready for the challenges they bring.

There's No Doubt that Big Data is a Big Deal to Companies Today

The benefits of big data include greater insight into customer sentiment, improved employee productivity, smoother operations and processes, and better decision making. And it's not just talk; a growing number of companies are taking action to implement big data projects. According to a recent survey by IDG, 49 percent of the 751 respondents say they are implementing or are likely to implement big data projects in the future, with 12 percent reporting that they have already implemented such projects.

As big data projects move from the planning to implementation stage, however, many companies are learning that they aren't prepared for all of the changes that these projects bring. Big data by definition involves very large quantities of unstructured data in various formats that often change in real time. Because big data encompasses so much information in so many formats that must be pulled together for analysis, it has significant impact on enterprise networks and IT infrastructures. These range from resources and equipment to deployment and management. In addition, big data projects tend to be initiated high up in a company's hierarchy. As a result, network managers and IT administrators

are often left to deal with these issues once a project is already underway—in other words, once it's too late. This creates the potential for a worst-case scenario: A big data project fails to reach stated goals because the necessary IT resources aren't available to support it, and the project strains the organization's network and servers to the point that other operations and processes suffer.

"With big data projects, there's a lot of information coming in, and it has to be in real time in order to be effective and precise. But you can't be pulling in that much data and not expect it to have an overall impact on network performance," says Jim Rapoza, senior research analyst of network and application performance at Aberdeen Group. According to a survey conducted by Aberdeen that was published in July 2013, the impact of big data on enterprise networks was ranked as the No. 4 performance concern by respondents, the majority of whom said their organization had already implemented big data projects or planned to in the next 12 months.

Big Data Packs a Punch

The impact that big data has on enterprise networks and IT infrastructures is multidimensional and driven by the three Vs: volume (growing amounts of data), velocity (increasing speed in storing and reading data), and variability (growing number of data types and sources). In particular, big data can significantly stress and stretch the following network resources:

Bandwidth

Running big data analytics requires a lot of bandwidth on its own; the issue is magnified when big data and day-to-day application trafficare combined over an enterprise network.

Latency

The real or near real-time nature of big data demands a network architecture with consistent low latency to achieve optimal performance.

Capacity

Massive amounts of highly scalable storage are required to address the insatiable appetite of big data, yet these resources must be flexible enough to handle many different data formats and traffic loads.

Processing

Big data can add significant pressure on computational, memory, and storage systems, which, if not properly addressed, can negatively impact operational efficiency.

Secure Data Access

Big data projects combine sensitive information from many sources like customer transactions, GPS coordinates, video streams, and more. All of these must be protected from unauthorized access.

Each of the above requires consideration and evaluation before embarking on a big data project. All too often, the process of evaluating and understanding the capabilities of the enterprise network and IT infrastructure doesn't occur until a big data project is underway and a problem is encountered. "What often happens is people in the organization take the network

for granted, so they just expect that a big data project will work," says Brad Reinboldt, product marketing manager with Network Instruments. "When it doesn't, the business is impacted, and it's the network managers who people come looking for."

"48% of the respondents expect their network loads to double over the next two years, and 23% expect the value to triple during that time frame. Yet only 41% say they are ready for this surge in network traffic."

Source: Connected World Technology Report[1]

Key Steps to Prepare Networks for Big Data

In a recent Connected World Technology Report, Cisco revealed results of a survey of 1,800 IT professionals: Forty-eight percent of the respondents expect their network loads to double over the next two years, and 23 percent expect the value to triple during that time frame. Yet only 41 percent say they are ready for this surge in network traffic. These results show that network managers would be wise to get involved now and have a voice in big data conversations—even if you're not asked.

"You have to get a seat at the planning table and get involved to the extent that you're able," Reinboldt says. "For example, if the plan is to host a big data project in the cloud, network managers should get involved to make sure meaningful SLAs are in place."

[1] This data is from Cisco's survey that was conducted in August 2012, the most recent survey that the company reported on the topic at press time.

Network managers should take the following steps to prepare their networks and infrastructures for big data:

Evaluate Your Monitoring Strategy/Devices

Determine if the current plans, policies, and assets can handle the increased stress of big data. Answer these questions:

If the infrastructure is virtualized as part of a big data initiative, how does this affect your ability to monitor performance and resources?

- Are your monitoring tools designed to handle high network traffic speeds, including the ability to capture and write to disk for retrospective analysis capabilities?

- Can you create dashboards and long-term reports that provide the real-time and historical perspectives needed to successfully monitor the high-demand, high-availability resources and services involved in big data operations?

- If big data resources are deployed, are you tracking availability and response times from the cloud provider? Can your monitoring solution validate performance and SLA compliance?

Predeployment Assessment

Before rolling out big data, you've got to know your current application and infrastructure status. Then begin quantifying the implications of big data on resources with these questions:

- Obtain a baseline of all operational metrics across the infrastructure, services, and applications. This is critical to ensure that existing resources do not suffer degraded performance after the rollout. Do you have this information at the ready for post–big data deployment evaluation?

- With big data's stiff latency and bandwidth demands, making the right deployment decisions takes on added urgency. Achieving the optimal execution of these initiatives means architecting for speed and quickness. Are all IT infrastructure teams working toward this objective?

- Will the data be stored and processed locally or in the cloud? Are your service providers on board to consistently deliver the required performance and uptime? Make it official with SLAs to achieve reliable levels of performance.

- Is your network configured to minimize latency and maximize bandwidth in the transport and processing of big data?

- Who needs the data? How will users be accessing it (wired or wireless connections), and does that impact performance?

Gain Contextual Awareness

Big data solutions will most likely run in a diverse vendor application environment. Therefore, network managers need to understand and anticipate big data's impact on:

- Server resource management—Given the computationally intensive nature of big data analytics on processing loads and storage performance, quantifying underlying health is a must. For example, it's important to track utilization metrics of virtual machines (VMs) running alongside the underlying host utilization metrics in order to ensure that adequate resources are allocated to support any VMs that would be involved with a big data application. This has direct implications on network behavior.

"People in the organization take the network for granted, so they just expect that a big data project will work. When it doesn't, the business is impacted, and it's the network managers who people come looking for."

– Brad Reinboldt
Product Marketing Manager
Network Instruments

- Applications—Because of its size and velocity, big data magnifies the importance of understanding all applications, their architecture, where they are hosted, and the location of end users, along with the underlying resources to support them and how they interact with each other on the network. Why is this important? Remember, all your services and applications run over the same network pipes, commingling and consuming resources. To this already volatile mix, you are now adding big data, an application on steroids.

- Cloud deployments—When WANs are involved, there's less visibility and control over network performance. For example, calls to back-end databases can create more latency than expected, and that negatively affects the end-user experience. It's important to understand each tier and the implications if you choose to host them in a different way.

- Cross-silo correlation—With big data comes the need to relate information from disparate IT resource groups spanning the organization. Network managers must strive to quantify these resource interdependencies for reaching peak big data performance.

"With big data, we are in the midst of perhaps the most dynamic computing environment ever; there are massive changes going on right now."

–Aberdeen Group's Rapoza

Keep Big Data Projects Running Smoothly Once Implemented.

As with any large-scale IT initiative, planning, preparing, and deploying big data is only the first step. Keeping it—and the myriad other services—running flawlessly entails an ongoing effort. The good news is that you don't need to be a big data expert to fully understand and quantify its health. All stakeholders must collaborate to maintain an awareness of this dynamically changing IT service. With this in place, network teams can preserve their end of the deal with these practical steps:

- Capture and analyze traffic among all critical network devices—including within virtualized systems—to enable optimal performance.

- Monitor system and infrastructure performance metrics, and alert on high utilization of operational metrics such as I/O, CPU, memory, and disk access, since these types of variables can act as the "canary in the coal mine" to underlying resource stress.

- Scrutinize application performance (layers 5 to 7) at the host and from all client perspectives, whether mobile, thin client, or desktop as ultimately users interact with services here—where the rubber hits the pavement and all complaints begin.

- Leverage top-down dashboards across domains for big data performance correlations.

- Quantify big data application tier performance, since these solutions demand tuned, tightly coupled, and efficient interactions to achieve top operation efficiency.

And finally, it's important that network managers do their homework by reading, researching, talking to vendors and peers, and staying up to date on big data trends and other technologies that affect big data.

"With big data, we are in the midst of perhaps the most dynamic computing environment ever; there are massive changes going on right now," syas Aberdeen Group's Rapoza. "The way networks perform is going to be completely different in five years' time. You have

to do the research and the reading to understand how technology changes, and build networks that are ready for those changes."

The Right Tools for the Job

Taking the necessary steps to prepare your IT resources for big data projects and keeping them running optimally is best done with the right tools in hand.

Performance monitoring tools help you see the big picture by providing visibility into network, infrastructure, and applications—while also delivering high-level organizational awareness. Ideally, they should either support high network speeds, or rely on a network monitoring switch/network packet broker to intelligently manage traffic to the monitoring devices.

These tools help you capture and manage traffic to ensure accurate monitoring of critical applications. Such tools also help you to monitor and understand how your cloud-based deployments impact your in-house IT resources. By using these tools to achieve a baseline of your network's capabilities and vulnerabilities now, you'll be better prepared to determine what investments and upgrades might be required to help make big data projects a success—and ultimately enable your business to thrive.

Once big data is deployed, monitoring solutions help you track the performance of big data traffic on network pipes and help IT infrastructures perform at their best. And when the network resources are not reaching performance goals, these solutions enable identification of the problem and quick resolution to maintain a higher standard of end-user satisfaction and business results.

Worth It in the End

There's a lot at stake with big data. Big data rollouts can be costly, and the pressure is on to show tangible value from these projects. After all, the upside to big data can't be denied. Projects that are thoughtfully implemented can deliver unprecedented value to the business, offering insight, productivity, and innovation—a recipe for success.

Get the Complete Picture

Network Instruments' integrated suite of solutions provides enterprise-wide application and network performance monitoring. With layer 2 to layer 7 visibility, these solutions offer real-time dashboards, comprehensive reporting, expert analytics, infrastructure awareness, and high-speed packet capture for retrospective analysis. Network Instruments' offerings help you proactively monitor IT resources, and when a problem arises you can quickly triage, troubleshoot, and resolve it to minimize the impact to the end user. When you get the complete picture, keeping your company's operations running smoothly is a cinch.