ACHIEVING THE DYNAMIC DATA CENTER

THE ROLE OF CLUSTERED INFRASTRUCTURE IN MANAGING EMERGING WORKLOADS

Brought to you by:

Bright Computing

Table of Contents

Chapter 1

Executive Summary

Today, three important conversations are taking place inside of the enterprise: How do we accelerate innovation in our products and solutions to serve our markets better? How do we leverage data to increase insight about our customers, our markets, and our operations? Finally, how do we establish greater agility to take advantage of new ideas and new opportunities that are coming to us?

As organizations strive to achieve greater innovation, insight, and agility, IT is under pressure to change. To change faster, in fact, than ever before. New business imperatives are driving the need to execute an emerging set of computational and analytical workloads with very different characteristics from traditional IT.

With so many new workload types emerging in the enterprise data center, IT teams are faced with the rapid adoption of technologies ranging from high performance computing and big data analytics to virtualization, containerization, and cloud. And of course, all of this must be achieved while keeping IT budgets reined in.

While this may seem daunting, IT teams can benefit from a unifying thread that connects these seemingly diverse and disparate technologies: all of them, in one way or another, are based on the common principle of clustering. Clustered systems running the Linux operating system, specifically, are coming to the foreground as the de facto standard for building out standalone -- or better yet integrated -- capabilities in each of these emerging areas.

Linux-clustered systems, while incredibly powerful, are also complex to deploy, manage, and monitor. IT teams facing the challenge of deploying Linux-clustered systems must move quickly to implement effective management disciplines across them.

This white paper examines the emergence of clustering in the enterprise data center. Subsequently, this document proposes a set of criteria to consider when evaluating options for managing advanced, clustered IT infrastructure in today's modern, dynamic data centers.

Chapter 2

Challenges of the Modern Data Center

The Business Imperative: Innovation, Insight, and Agility

Three important conversations taking place inside the enterprise:

How do we accelerate innovation in our products, solutions, and services to address the market better? Organizations today have widening access to technology that can help tailor the design and production of goods and services to the ravenous and ever-changing appetites of consumers. Computer-aided design and engineering, computer-based simulation and modeling techniques, and 3D printing are but a few important examples of technology that is revolutionizing the way things are made. Broad availability of these new tools is upping the competitive ante, suggesting that the organizations best equipped to harness them will reap the greatest market rewards.

How do we leverage data to increase insight about our customers, our competitors, or our operations? Across every industry, organizations face rapidly shifting market, customer, and competitive landscapes. The pace of change threatens to outstrip the decision-making speed and capacity of even the best-executing teams. This dynamic explains the rise and rapid embrace of computing disciplines that promise to distill meaningful insights from raw and unstructured business data, such as big data analytics, machine and deep learning, and the "internet of things." All of these disciplines are predicated, however, on the ability to build and maintain computing platforms capable of storing, processing, and analyzing unprecedented volumes of data.

How do we establish greater agility to take advantage of new ideas and new opportunities that are coming to us? With so much dependency on technology to drive innovation and insight, organizations can only achieve business agility if underpinned by a corresponding focus on IT agility. Organizations must develop sustainable competencies around selecting, implementing, and managing new technologies that can support emerging, business-critical workloads. Importantly, these technologies must be deployed across an IT infrastructure (whether on-premise, outsourced, or a combination) that is both extremely reliable and extremely flexible.

IT Transformation: New Workloads Driving Unprecedented Change

As organizations strive for greater innovation, insight, and agility, IT is under pressure to transform itself and align to these business imperatives. Emerging, business-critical workloads being deployed by more and more organizations include, but are not limited to:

- Data analytics and sequencing
- Machine learning / deep learning
- Real-time modeling and simulation
- Rapid application development and deployment
- Computer-aided design (CAD) and engineering (CAE)

Businesses that are turning to advanced technologies like these -- whether to find their unique competitive edge, predict the next must-have customer trend, or unearth the latest scientific revelation -- find their fortunes resting on IT's ability to run these increasingly sophisticated computational and analytical workloads in a more agile and dynamic way.

IT's Dilemma: "In With the Old, In With the New..."

No matter how much IT wants to step up and align to the business, the reality is that most enterprise data centers are not currently optimized to handle these highly varied, complex, and dynamic workloads. In fact, the introduction of these new workloads may cause a bit of organizational "whiplash" within enterprise IT teams. This is because IT is accustomed to being asked for -- and has made tremendous strides in delivering -- infrastructure with attributes like these: Stable, Secure, Predictable, Reliable, Efficient, Automated.

As expected, no one in the line of business is saying that these hard-won attributes should now go away. Rather, the assumption is that IT will continue to deliver stable, secure, and reliable services while simultaneously supporting the new types of workloads required to drive innovation, insight, and agility. These new workloads, in turn, demand IT infrastructure with attributes like these: Flexible, Scalable, Resizable, Virtualized, Containerized, Cloud-optimized, Software-defined.

How does enterprise IT take decisive steps toward creating more agile, flexible IT infrastructure for its dynamic data center? And importantly, how does IT achieve this without sacrificing recent gains in terms of security, reliability, and efficiency?

TRADITIONAL

- Stable
- Secure
- Predictable
- Reliable
- Efficient
- Automated

ADDITIONAL

- Flexible
- Scalable
- Resizable
- Virtualized
- Containerized
- Cloud-optimized
- Software-defined

Chapter 3

A Path Forward: Achieving the Dynamic Data Center

Leading edge IT organizations are finding a path forward, toward agile infrastructure that balances secure and reliable IT service delivery with flexible and dynamic support for emerging workloads. In fact, more than one path is available; enterprise IT might opt to run separate IT infrastructures optimized for different types of workloads, or alternatively might decide to combine the best of all attributes into a single, blended IT infrastructure. Let's explore this in a bit more detail.

Gartner has coined the term "bimodal IT" in reference to enterprise organizations that opt to run two IT infrastructures in parallel, one optimized for traditional IT workloads and a second, separate infrastructure geared more dynamically to handle new workloads. According to Gartner, bimodal IT is "the practice of managing two separate, coherent modes of IT delivery, one focused on stability and the other on agility." [1] These parallel infrastructures might co-exist side by side within a single data center or -- alternatively -- might be housed and managed in completely different parts of the enterprise.

Not every IT organization, however, believes that a bimodal approach is viable or desirable. Many are striving instead for a single, blended model that combines the best attributes of both "traditional" and "dynamic" IT infrastructure. In a blended model, the same enterprise IT data center team and infrastructure are tasked with running any IT workload or service with equal competency, and with a balanced emphasis on stability and agility.

Either way, as these new workloads find their way into the data center, IT is left grappling with a broader range of technologies and capabilities than in the past. This introduces many new challenges in terms of skills, staffing, resource management, and service delivery, which we will return to shortly.

The Central Role of Clustering

One likely centerpiece of the data center's architectural makeover is clustering to address emerging workloads. Linux clusters, in particular, are prevalent in most data centers handling the kinds of workloads that require either massive real-time computations (think simulation and modeling) or massive data processing and analysis (think deep learning).

Why Linux clusters? Simply, because the computational and processing power that can be achieved when clustering Linux servers together is hard to achieve in any other way, and certainly cost-prohibitive using more traditional approaches. The industry blog Tecmint defines Linux clustering as "establishing connectivity among two or more servers in order to make it work like one" and goes on to explain that "clustering is a very popular technique among systems engineers who want to cluster servers as a failover system, a load balancing system, or a parallel processing unit." [2] Linux is also attractive as an open source platform with strong community, foundation, and commercial support.

Two of the most popular types of workloads to run on Linux clusters are high performance computing and big data analytics.

High performance computing (HPC) is defined by TechTarget as "the use of parallel processing for running advanced application programs efficiently, reliably and quickly." [3] Some specific examples of workloads performed by HPC systems include genetic sequencing, climate modeling, and financial modeling, all of which share the characteristic of requiring massive parallel computations to be executed in real-time.

IDC highlights the growing importance of HPC for today's commercial enterprise, stating that, "CIOs who do not assess the potential benefits of HPC will increasingly risk losing ground to competitors that have learned how to exploit HPC to accelerate and improve innovation." Further, a recent, joint study by IDC and the Council on Competitiveness found that, "97% of the commercial firms that had adopted HPC said they could no longer compete or survive without it. The chief benefit cited by these firms is that HPC enables them to bring more innovative, higher-quality products and services to market in shorter time frames." [4] These forces are driving HPC clusters to become pervasive across many industries.

Big data analytics is defined by Wikipedia as the ability to analyze "data sets that are so large or complex that traditional data processing applications are inadequate." [5] Today, commercial enterprises understand it is imperative to glean as much insight as possible from all available, relevant data sources. With every computer, mobile device, or sensor-equipped "thing" in the new Internet of Things (IoT) paradigm capable of serving up a nonstop data stream, however, the challenge of turning raw inputs into meaningful insights can quickly become overwhelming.

In response, open source technologies such as Hadoop and Spark have emerged to place powerful analytical capabilities in the hands of any-sized team or organization. Meanwhile, deep learning and machine learning have emerged as sophisticated variants of big data analytics, both promising to go even further, faster, and deeper in gleaning insight from massive data sets. As Stanford Professor and Coursera CEO Andrew Ng points out, "Machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome." [6] Linux clustering underpins these and many other pioneering initiatives that have successfully harnessed big data.

The Emergence of Hybrid Cluster Environments

Increasingly, organizations will identify both HPC and big data analytics as business-critical disciplines. In fact, they are highly complementary, with HPC bringing a real-time, theoretical, simulation and modeling-based dimension to decision-making, while big data analytics examines actual data sets gathered over an extended period of time. The ability to blend these two computing disciplines -- "big compute" and "big data" -- into a single, integrated decision support system has been coined high performance data analytics (HPDA) by Steve Conway of IDC. According to Conway:

> "Leading commercial companies in a variety of market segments are turning to HPC-born parallel and distributed computing technologies — clusters, grids, and clouds — for challenging big data analytics workloads that enterprise IT technology alone cannot handle effectively." [7]

To achieve HPDA, organizations need the ability to run HPC clusters, Hadoop clusters, and adjacent technologies, such as Spark, across a common infrastructure. This may mean running separate clusters side by side, sharing a common storage layer. However, the data center's relentless drive for efficiency may instead dictate that the same computing resources need to be repurposed dynamically to execute different types of workloads -- compute or data intensive -- at different times.

Whether running HPC and big data clusters side by side, or provisioning the same compute resources dynamically to execute either type of workload, data center IT teams need tools to assist them in managing these complex computing resources easily, flexibly, and as centrally as possible.

Accelerating Agility: Virtualization, Containers, and Clouds

HPC and big data clusters aren't the only types of advanced, clustered infrastructure finding their way into the enterprise data center. As organizations -- and the IT teams that support them -- strive to be more agile in response to rapid-fire business requirements, this is also driving mainstream data center adoption of virtualization, containerization, and cloud computing. Whether implemented separately or in combination, each of these offers IT the potential to use resources more efficiently, while also delivering services more responsively.

Interestingly, the widening use of virtualization, containers, and cloud computing contributes to the adoption of Linux-based clusters in the enterprise data center. Linux-based Kubernetes clusters sit at the core of containerized applications, while private clouds – such as an OpenStack cloud -- are also typically implemented atop Linux-based clusters.

Moreover, both private and public clouds increasingly serve as the underlying infrastructure layer upon which compute clusters – such as HPC or big data clusters – are built. The ability to spin up a virtualized HPC or big data cluster within minutes, tailored to a user's specified level of computing needs and length of required access, can deliver massive improvements in IT resource efficiency and user satisfaction.

As big compute, big data, virtualization, containers, and clouds invade the enterprise data center, a common thread that runs through all of these – even if not immediately apparent to all stakeholders – is Linux-based clustering.

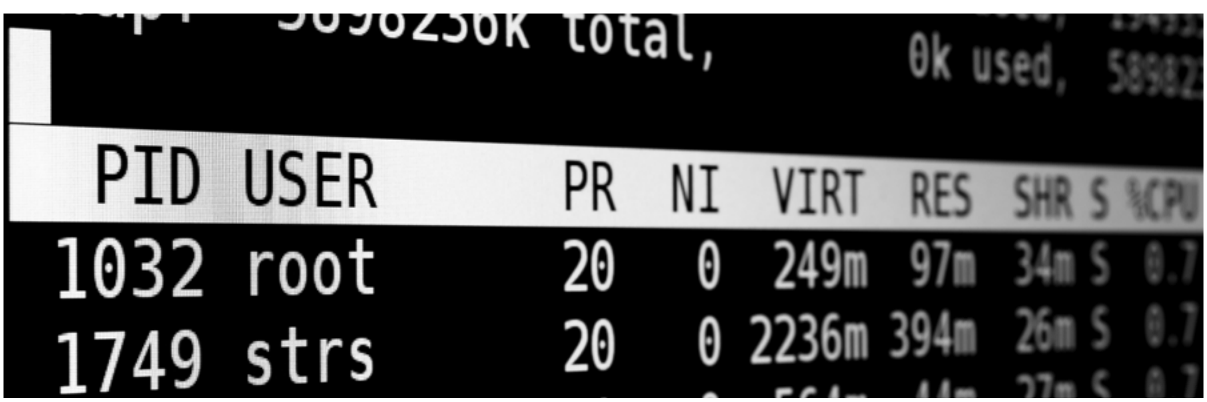
Implications of Linux Clustering for the Enterprise Data Center

In many organizations, Linux-based clusters have initially emerged outside of the central data center, deployed and attached to standalone high performance computing, big data analytics, or other exploratory line of business initiatives. As these initiatives mature and become more central to the business, however, the centralized IT team is typically enlisted to bring some or all of the clustered resources under centralized management.

This poses a daunting challenge to IT teams already drowning in business applications, diverse workloads, and highly varied systems and platforms. While incredibly powerful and versatile, Linux clusters can also be very challenging to set up and to manage over their entire lifecycle. IT teams accustomed to supporting traditional workloads and applications -- which run on non-clustered systems -- may lack the skills, bandwidth, and operational procedures required to absorb clustered IT infrastructure into the data center.

These resource barriers can inhibit or delay the deployment of clusters, preventing organizations from running the emerging workloads that are increasingly essential to the success and growth of the business. Further, once deployed, the clustered infrastructure may pose ongoing challenges and operational risks to IT teams lacking requisite skills, resources and processes.

With so many business-critical workloads, applications, and systems converging around this powerful notion of clustering, how will IT cope?



Chapter 4

Managing the Dynamic Data Center

Converged, Clustered IT Infrastructure Demands A New Management Paradigm

Without question, IT teams facing the challenge of absorbing Linux-clustered systems must quickly identify effective and sustainable approaches to deploying, managing, and monitoring clusters. As might be expected, some approaches work better than others. The most commonly adopted paths to managing clustered systems are:

Scripting and open source tools. Frequently, when tasked with deployed and administering a Linux-based cluster for the first time, a resourceful and talented IT team will rely initially on "low-cost" or "free" approaches such as downloading open source tools and writing homegrown scripts to manage various components of the clustered infrastructure. However, this quickly leads to a "cobbled together" platform that can break down very quickly, since changes in one part of the platform are not propagated seamlessly to others. Compounding this fragility is the fact that knowledge becomes codified in individual users, coders, and scripters, rather than in the platform itself. This introduces unwelcome risk and exposure in the face of inevitable personnel changes over time.

Hardware-specific management tools. Some providers of commodity servers package their own management tools along with clustered hardware. These vendor-specific tools should be closely examined for their breadth of software functionality. Another consideration is that vendor-specific tools, while potentially suitable for homogenous systems, are by definition inadequate to stretch across multi-vendor clustered infrastructure.

Outsourced management. Some data centers may shy away from bringing clustered infrastructure on premise, opting instead to explore outsourcing of clustered workloads and platforms to managed service or cloud providers. In these scenarios, the cluster management burden shifts to the outsourced provider of compute resources. Organizations evaluating outsourced approaches to clustering must consider factors including cost, control, performance, latency, and security before determining the optimal blend of outsourced versus on premise clustering.

Enterprise-class management platform. True "enterprise-class" management of clustered infrastructure requires characteristics lacking in the alternative approaches outlined above. Commercial enterprises need an approach to managing clustered infrastructure that balances flexible and agile execution of business-critical workloads with efficient and effective utilization of IT resources. Over time, most organizations determine that a vendor-independent, commercial-grade management platform is required to achieve this optimal balance.

For organizations ready to implement enterprise-class management of their business-critical, clustered workloads, what criteria should be considered in evaluating available platforms?

Enterprise-Class Cluster Management: Essential Criteria

A truly enterprise-class cluster management platform can serve as a powerful control point across many different types of clustered IT infrastructure. It provides services in a user-friendly way that enables IT administrators to manage complex, clustered infrastructure without acquiring additional, specialized skills. To support administrators in deploying, managing, and monitoring clustered infrastructure effectively, the management platform should meet these essential criteria:

Single pane of glass. As more and more technologies, applications, and workloads migrate into the enterprise data center, if left unchecked, systems administrators will be forced to contend with a multitude of management interfaces. Adopting a unified management platform that provides "single pane of glass" visibility and insight across all clustered IT infrastructure eliminates the need to manage each clustered system or workload as a standalone entity. With the aid of this single, powerful console, systems administrators can deploy, manage, and monitor all types of clusters to meet rapidly changing business and resource requirements.

Ease of deployment. Turning racks of commodity servers into a fully functioning cluster can be a monumental task when executed using homegrown scripts or open source tools. The process can be lengthy and cumbersome, opportunities for error abound, and the challenge repeats itself every time a new cluster must be spun up or an existing cluster expanded, modified, or redeployed. An enterprise-grade cluster management platform, alternatively, empowers IT administrators with robust, repeatable and highly automated set-up procedures. An entire cluster can be fully functional in as little as an hour, and the risk of misconfiguration greatly reduced in the process.

Lifecycle management. Once clustered infrastructure is deployed, it rarely remains static. Managing a cluster through its lifecycle entails constant provisioning and configuration updates to address changing workloads, as well as changes to the underlying physical infrastructure. Typically, the same updates must be executed across every server or "node" on a cluster, an extremely time-intensive task when administered manually. An effective cluster management platform removes this time burden by orchestrating updates via a central "head node" and propagating these changes out in a consistent and automated way to every node on the cluster, even when the cluster count is in the thousands.

Scale out, scale in. One of the most important changes that must be managed in any cluster is adjusting the allocated resources to address the workload task(s) at hand. To minimize physical resource consumption, and to conserve energy and other resources, most enterprise data centers want to allocate only the minimum required compute resources to any specific workload, switching other nodes off or reallocating them to address alternative workloads in parallel. The flip side of this "scale in" requirement is the need to add nodes dynamically to a cluster when workloads demand additional compute resources. Adopting an enterprise-class cluster management platform enables IT administrators to dynamically allocate nodes as required to support workloads of all sizes and types.

Monitoring and healthchecking. When clustered infrastructure is running mission-critical workloads on behalf of a line of business, cluster downtime and poor cluster performance can't be tolerated. Like any IT infrastructure that is integral to the front lines of the business, clusters must be monitored comprehensively to ensure that all availability, stability, security, and performance thresholds are being met. Furthermore, most data centers will want to configure and execute automated responses to known issues, so that IT resources aren't tied up executing predictable, manual remediation tasks. Enterprise-class cluster management must empower IT to respond intelligently and automatically to issues identified via monitoring and healthchecking.

Breadth of integration. Any management platform's value-add to an enterprise data center is closely tied to the breadth of technologies it is equipped to detect, manage, and monitor "out of the box." In general, a management platform that can provide insight across a broad cross-section of the data center should be favored over point solutions that handle only narrow technology niches. Linux clusters come with a whole host of new technologies that must be managed, such as the clustered server hardware, the operating system, GPUs or accelerators, CPUs, interconnects, workload schedulers, programming environments, etc. Enterprise-class cluster management must provide out of the box integration with the broadest possible range of hardware vendors, software vendors, packages, tools, and files deployed in Linux clusters.

Enterprise-Class Cluster Management: Advanced Criteria

As the enterprise data center embraces more advanced capabilities to extend the power of clustered infrastructure, the cluster management platform must also be evaluated against some or all of these criteria:

Hybrid cluster management. Increasingly, organizations will require composite decision support systems that support both real-time (big compute) and trended (big data) types of analysis. Such a system can be architected by deploying a combination of high performance computing (HPC) and big data (Hadoop) clusters. IDC refers to this combined set of capabilities as high performance data analysis or HPDA. Further, the relentless drive for data center efficiency may dictate that different cluster types must be deployed across the same infrastructure at different times. IT teams must be able to dynamically reprovision cluster nodes to support varied workloads.

CPU, GPU, accelerator, and interconnect management. Central processing units (CPUs), graphics processing units (GPUs), accelerators (such as Intel's MIC technologies), and low-latency interconnects are all commonly deployed elements of high performance Linux clusters and must be comprehensively managed, monitored, and health-checked accordingly. Plug-and-play integration with the most common CPU, GPU, accelerator, and interconnect platforms is essential for IT teams aiming to maximize cluster compute power and performance.

Public cloud bursting. Since clustered workloads tend to vary widely in terms of size, duration and required compute power, it can be next to impossible for IT teams to build appropriately sized clusters that optimize utilization at all times. An attractive alternative is to right-size the on-premise, clustered environment to address a predictable level of run-rate activity, and to "burst" to the public cloud when demand for clustered resources spikes. An enterprise-class management platform must therefore support two specific capabilities: First, IT should be empowered to expand an existing cluster dynamically by accessing incremental nodes in the public cloud. Second, IT needs the ability to spin up an entirely new cluster in the public cloud, with a desired, pre-defined set of attributes. Finally, organizations may opt to send less critical or commercially sensitive clustered workloads to the cloud as a matter of routine.

Private cloud deployment. While some IT tasks are ideally suited for the public cloud, enterprise IT data centers are also embracing private cloud for a host of reasons, ranging from cost and control to performance and privacy. OpenStack, an open-source framework for private cloud deployment, has gained considerable popularity among data centers that are adopting private cloud and looking for a cost-effective alternative to traditional private cloud platforms. Today, OpenStack-based private clouds are being widely deployed for development, test, and even production-level applications. Since OpenStack is exclusively deployed on Linux-based clusters, an enterprise-class cluster management platform must encompass the ability to deploy, manage, and monitor OpenStack private clouds.

Virtualized clustering / cluster-as-a-service. One reason a data center may wish to deploy an OpenStack-based private cloud is to run clustered applications and workloads inside of it. Just as a private cloud can streamline the operation and utilization of IT infrastructure for traditional IT workloads, it can do the same for the more complex, emerging workloads that need to be run on clustered systems. Setting up and running virtualized clusters, inside of an OpenStack private cloud, can offer significant efficiency by right-sizing the physical hardware footprint needed to deploy and run clusters. A truly enterprise-class cluster management platform will enable virtualized clusters to be set up, managed, and monitored in all the same ways as physical clusters, along with the added efficiency benefits of being deployed virtually.

Container support. No solution for managing the modern data center would be complete without addressing containers. They serve the same purpose as virtual machines, but have the advantage of providing bare-metal performance. This is because only the application software is virtualized, while the underlying operating system and services is shared among them. Containers are being used to build distributed applications, provide service resilience, and data center portability. Setting up and maintaining them can be a challenge without a capable clustered infrastructure manager.

Appendix

Bright Computing: A Powerful Management Platform for Clustered IT Infrastructure

Bright Computing is the world's leading independent provider of enterprise-class management software for clustered IT infrastructure, including HPC clusters, big data clusters, and OpenStack private clouds. Bright software is designed to scale; some Bright-managed clusters contain thousands of nodes. More than 500 organizations worldwide have selected Bright Computing to deploy, manage, and monitor their business-critical IT infrastructure. Bright's customer base includes global academic, governmental, financial, healthcare, manufacturing, energy, and pharmaceutical organizations such as 3M, Boeing, Cambridge University, Chevron, Cisco, CSIRO, ING Bank, Merck, NASA, PACCAR, Roche, St. Jude Children's Research Hospital, Saudi Aramco, Stanford University, Toyota and Volvo. Bright partners with Amazon, Cray, Dell, Intel, Mellanox, NVIDIA, SGI, and other leading vendors to deliver powerful, integrated solutions for managing today's most advanced, clustered systems.

To learn more, request a demo, or initiate a free product evaluation, contact Bright at www.brightcomputing.com.

References

1. Gartner IT Glossary : Bimodal IT, Gartner
2. Introduction and Advantages/Disadvantages of Clustering in Linux, Tecmint.com
3. Definition: high-performance computing (HPC), TechTarget.com
4. "10 Things CIOs Need to Know About High-Performance Computing," Steve Conway, IDC, March 2016.
5. Big Data, Wikipedia.org
6. Machine Learning, Coursera.org
7. "High Performance Data Analysis: When Big Data Needs More Firepower," Steve Conway, IDC, June 2014.