DataCenter Knowledge™

THE STATE OF CLOUD

A look at OpenStack, Automation, and Managing Data

White paper sponsored by:

Author: Bill Kleyman

DELL EMC

intel

DataCenter Knowledge

CONTENTS

Introduction:

Cloud computing, or simple "cloud," has become a widely accepted model and technology platform for delivering a wide range of IT services in a highly efficient manner. In fact, it has become nearly ubiquitous, with accelerating rates of adoption across all types of businesses. However, the idea behind the cloud isn't an all-or-nothing mentality. Really, it's quite the opposite. Cloud computing provides great flexibility to the organizations that deploy it right. In some cases, this means utilizing the entire cloud solution; while in other instances, only a part of the cloud service is necessary.

Either way, growth around cloud will only continue to increase. Specifically, IT spending is steadily shifting from traditional IT offerings to cloud services, known as "cloud shift," according to Gartner. The aggregate amount of cloud shift in 2016 rose to $111 billion and is projected to increase to $216 billion in 2020.

Furthermore, Gartner analysts said that by 2020, cloud, hosting and traditional infrastructure services will come in more or less at par in terms of spending.

"As the demand for agility and flexibility grows, organizations will shift toward more industrialized, less-tailored options," DD Mishra, research director at Gartner, said in a press release. "Organizations that adopt hybrid infrastructure will optimize costs and increase efficiency. However, it increases the complexity of selecting the right toolset to deliver end-to-end services in a multi-sourced environment."

Gartner predicts that by 2020, 90 percent of organizations will adopt hybrid infrastructure management capabilities.

So, there is no question that IT and business are inextricably linked. Ever since the internet revolutionized the business world in the 90's, companies around the globe have turned to technology to increase business efficiency while boosting their competitive edge. The link

between business and IT has created unprecedented growth and opportunity across all industries, making IT a key discussion topic in the boardroom of most companies.

And, as a major part of that discussion, cloud services are a huge consideration when deploying next-generation applications, working with big data, and utilizing new types of cloud management solutions.

In this whitepaper, we're going to explore the current state of the cloud, ways that organizations are managing various cloud environments (using OpenStack, for example), and how they optimize the delivery of critical data points. Furthermore, we'll cover key management points as well as automation best practices.

The cloud will be distributed with 60% of IT done off-premise and 85% by multi-cloud by 2018.

Section 1: Cloud, Services, and New Requirements

A recent WSJ article points out that CIOs are knitting together a new IT architecture that comprises the latest in public cloud services with the best of their own private data centers and partially shared tech resources. Demand for the so-called hybrid cloud is growing at a compound rate of 27%, which outpaces overall IT market growth, according to researcher MarketsandMarkets.

Here's the big factor to consider: The cloud will be distributed with 60% of IT done off-premise and 85% by multi-cloud by 2018.

But how do you manage all of this? What does a multi-cloud ecosystem actually look like?

- **OpenStack Overview and Definition.** According to the OpenStack Foundation, which manages a lot of the progress around OpenStack, this platform is a set of software tools for building and managing cloud computing platforms for public and private clouds. Backed by some of the biggest companies in software development and hosting, as well as thousands of individual community members, many see OpenStack as the future of cloud computing, automation, and management. OpenStack lets users deploy virtual machines and other services that handle different tasks for managing a cloud environment on the fly. Furthermore, this type of platform helps support multi-cloud architectures for powerful scalability, resiliency, and business enablement. For example, a mobile app that needs to communicate with a remote server might be able to divide the work of communicating with each user across many different instances, all communicating with one another but scaling quickly and easily as the application gains more users.

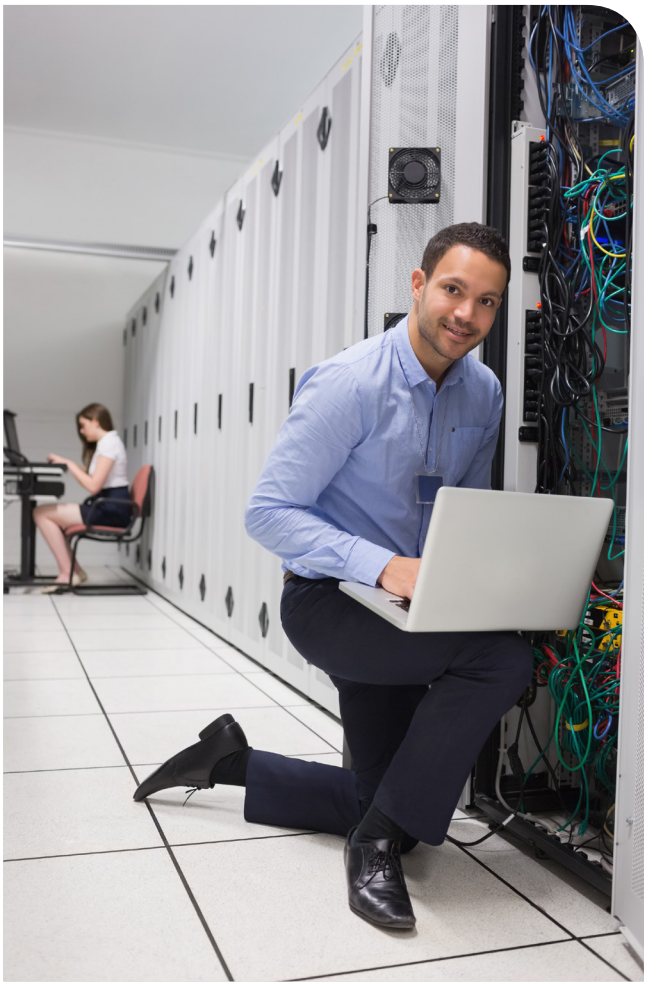
Many see OpenStack as the future of cloud computing, automation, and management.

- **Current State of Cloud Automation.** Essential to cloud automation is having a computing platform that has extensive and complete libraries of Application Programming Interfaces (APIs) that expose all the functionality of the cloud in a way that can be managed and controlled via automation software. Today, organizations are considering OpenStack technologies to drive their transformation initiatives because of its ability to foster innovation. OpenStack provides a platform for the rapid development, testing, and development of mission-critical and highly scalable applications. To create this level of rapid innovation, cloud automation is being used as a powerful tool to empower the business. A recent report from VMware and IDC indicated that the demand for datacenter and cloud automation software and SaaS solutions continues to expand as more organizations recognize that modern application architectures and multi-cloud infrastructure strategies require IT operations teams to be able to rapidly provision, update, and scale distributed applications, as well as physical, virtual, container, and cloud resources to support digital transformation initiatives.

- **The Importance and Growth Around Data.** We, as users and business owners, are creating so much data. The important point here

is that this data isn't benign. It's valuable, can help create market insight, and can help organizations create real-world competitive advantages. Gartner recently pointed out that data and analytics will drive modern business operations, and not simply reflect their performance.

Furthermore, executives will make data and analytics part of the business strategy, which will allow data and analytics professionals to assume new roles and create business growth. Shifting the way organization uses data and analytics more toward driving business operations requires a new approach to data architectures, which many organizations are already building. Last year, Gartner research found that 45% of IT professionals had indicated that new data and analytics projects were in the "design" and "select" phases. New solutions like those around Hadoop in an OpenStack cloud, provide optimized configurations based on proven architectures. These configurations lessen the need to tune your hardware and cloud software to enable the cloud environment to best meet the demands of your Hadoop workloads. But, it doesn't end there. Our reliance on data will only continue to grow.



A new update to IDC's Big Data and Analytics Guide now forecasts worldwide revenues for big data and business analytics (BDA) will reach $150.8 billion in 2017, an increase of 12.4% over 2016. Commercial purchases of BDA-related hardware, software, and services are expected to maintain a compound annual growth rate (CAGR) of 11.9% through 2020 when revenues will be more than $210 billion.

Moving forward, data analytics, working with multi-cloud management platforms like OpenStack, automation, and data center integration will become the norm. "We are at an inflection point as digital transformation efforts shift from 'project' or 'initiative' status to strategic business imperative," Frank Gens,

Senior Vice President and Chief Analyst at IDC, said in a press release. "Every (growing) enterprise, regardless of age or industry, must become 'digital native' in the way its executives and employees think, what they produce, and how they operate."

Section 2: Core Components of a Modern MultiCloud Platform

Administrators are leveraging powerful tools like OpenStack and automation to manage cloud, big data, and so much more.

With all of this discussion around digital transformation, we need to understand the critical role that automation and management play. When you're working with a multi-cloud system, you're not only integrating with underlying virtualization, you're also designing a more efficient data center management platform. You're combining network, storage, and compute into the management layer to control your most critical resources and optimize your users. The beauty of the modern cloud and data center architecture is that you can create intelligent network and management policies which scale on premise systems and into the cloud. This kind of seamless cloud delivery allows the user to be continuously productive while still access either on-premise or cloud-based resources.

To accomplish this, administrators are leveraging powerful tools like OpenStack and automation to manage cloud, big data, and so much more.

- **OpenStack and Automation.** A major component around multi-cloud architecture is the ability to manage and automate the platform. This is where powerful tools connect cloud, data center infrastructure, and automation into a complete solution. For example, a recent announcement for the next generation of the Dell EMC Ready Bundle for Red Hat OpenStack Platform showcased a production-grade cloud platform which was jointly engineered and validated by Dell EMC and Red Hat. This solution features Red Hat OpenStack Platform (OSP) 10, a massively-scalable and agile cloud Infrastructure-as-a-Service (IaaS) solution based on the upstream OpenStack 'Newton' release. The Ready Bundle for Red Hat OpenStack delivers new and enhanced capabilities not only in features but also in ordering, deployment, and support capabilities, as well. The JetPack Automation Toolkit allows for the rapid and automated deployment experience that's been refined and optimized over the years. With this release, you're leveraging a single package for configuring the underlying hardware and the

OpenStack software in a fully automated fashion.

JetPack includes a large set of automation scripts that are used in conjunction with Red Hat OSP Director (based on the OpenStack OOO project) and OpenStack Ironic bare metal provisioning to provide a completely automated software deployment experience, as well as an automated update and upgrade experience

- **Leveraging OpenStack for Big Data.** We mentioned the importance of data. And we've discussed how new technologies around cloud management and automation are impacting business strategies. Now, organizations are combining big data requirements with OpenStack and automation platforms. Consider this use-case: Founded in 1958, Monash University was established to create a science and technology-focused research institution. Almost 60 years later, Monash is Australia's largest university, and now includes a curriculum that spans multiple disciplines and across five international campuses. The University's eResearch Centre fosters international collaboration and advanced research of both fundamental and applied sciences, with an emphasis on imaging and data science.

Where they got stuck: Eventually, the critical need to manage a continually growing pool of research data collided with the prohibitively huge investment presented by its previously deployed proprietary hardware solutions, leaving

the University at a crossroads.

How a new approach to data helped: Monash University selected Red Hat to implement a software-defined solution using Red Hat Ceph Storage on Dell EMC PowerEdge R630 and R730xd rack servers that can accelerate application performance, simplify systems management and address workloads at any level. Using Red Hat Ceph Storage as an object store, Monash projects that its software-defined storage solution will significantly reduce costs and give researchers the freedom to independently and incrementally add capacity and performance for future growth. Plus, when coupled with Intel architecture-based advanced analytics solutions, you can efficiently and effectively capture, process, analyze, and store vast amounts of data of all types. Built in partnership with industry leaders like Dell EMC, Intel's highly available, performance-optimized, open-standards-based solutions will support the most ambitious analytics-driven initiatives.

Research amplified and optimized: Monash University's eResearch Centre is now able to store and manage massive workloads of data, already encompassing five petabytes, within a single infrastructure. In addition to tight integration with Dell EMC, OpenStack components and services, and strong life cycle management, Red Hat Ceph Storage allows for replication and erasure coding capabilities for increased data protection and availability.

Monash Unversity's eResearch Center is now able to store and manage massive workloads of data, already encompassing five petabytes, within a single infrastructure.

"As a research institution, we are faced with the challenge of virtually limitless data, not only from new projects but from archived and long-tail research," said Steve Quenette, deputy director, Monash eResearch Center, Monash University. "One of our key concerns in this process was having enough storage space in an OpenStack cloud environment, as it supports the majority of use cases from our researchers. With Red Hat Ceph Storage (built on Dell EMC rack servers), we have been able to alleviate those concerns and have confidence in our ability to support future workloads with an intuitive, cost effective solution."

The system builds on Dell EMC's varied work in the high-performance computing space, with the Pearcey system installed in 2016 and numerous other systems for Australian universities such as the University of Melbourne 'Spartan', Commonwealth Scientific and Industrial Research Organization (CSIRO)

'Bracewell,' and the University of Sydney 'Artemis'.

- **Designing a Data Center Ready for Multi-Cloud.** The ability to integrate data center and cloud solutions with OpenStack architecture is a game-changer. Dell EMC has taken this strategy a step further and has formalized solutions that are designed to correspond to a customer's stage of OpenStack deployment. This means that whether you're just starting out with OpenStack, or have an established practice, Dell EMC is bringing these solutions to market via a "Build-to-buy" continuum with the goal of meeting their customers wherever they may land on this continuum. Dell EMC's "Build-to-Buy" offerings include a full range of solutions, from fully validated Ready Solutions, to reference architectures and system extensions, to advisory and implementation services. Some of the key offerings include:

Dell EMC Ready Bundle for Red Hat OpenStack Platform

Dell EMC NFV Ready Bundle for Red Hat

Optimized OpenStack solutions and reference architecture with Canonical, Mirantis, Red Hat, and SUSE

Partner plug-ins for ScaleIO, VNX, and XtremIO

Data protection extensions for OpenStack

OpenStack services portfolio from assessment to implementation

The core components of the future cloud and data center model must revolve around ease of use, simplified management, and rapid scale capabilities. Furthermore, we must be prepared for the world of multi-cloud. This is why working with solutions that are capable of integrating data center, cloud, and automation can help impact your ability to innovate at the pace of the digital market.

Section 3: NASA – A Real-World Use-Case

NASA's Goddard Space Flight Center in Greenbelt, Maryland, is home to one of the world's largest contingents of Earth scientists. These scientists investigate weather and climate phenomena at time scales ranging from days to centuries.

From analyzing historical weather and climate data to developing climate change projections for the decades to come, the center's researchers share a common need for high performance computing (HPC) systems built to handle massive datasets.

This is where the NASA Center for Climate Simulation (NCCS) enters the picture. A service-driven organization, the NCCS works to expand scientific and engineering frontiers by providing state-of-the-art supercomputing and data services for weather and climate researchers. It serves a broad user community based at NASA centers and laboratories and universities across the country and internationally.

The NCCS works to expand scientific and engineering frontiers by providing state-of-the-art super-computing and data services for weather and climate researchers.

- **The Challenge:** To carry out its mission, the NCCS operates an arsenal of powerful supercomputing resources. These include a massive and continually evolving HPC system, Discover, which currently encompasses around 90 racks and 90,000 total cores. While it was designed to power large-scale simulations, the Discover supercomputer was not optimized for the complexity of large-scale data analysis and applications that pair HPC capabilities with big data.

- **The Solution:** To meet their needs, and to alleviate some of the strain on Discover, the NCCS launched its Advanced Data Analytics Platform (ADAPT). The ADAPT system combines HPC and virtualization technologies in an onsite private cloud specifically designed for large-scale data analytics. In a novel and "green" twist, ADAPT was built largely from decommissioned HPC components that came out of the Discover supercomputer as that latter system evolved to bring in new technologies. This repurposed gear included hundreds of Dell EMC™ PowerEdge C6100 servers, which provide the bulk of the compute nodes in ADAPT, according to Garrison Vaughan, an NCCS systems engineer who was on the team that built ADAPT.

In a parallel effort, the NCCS is building a production OpenStack science cloud cluster within a physical container, separate from the NCCS

Climate Computing Facility's HPC environment. The OpenStack cluster in the modular data center will be a more flexible environment for various use cases, such as applications or operating systems not found on the HPC cluster, work patterns that are not conducive to a batch scheduler, and more inherently parallel, loosely coupled processing.

In addition, the OpenStack cloud will offer the potential for enhanced security. Virtual local area network (VLAN)-based OpenStack tenants will allow NCCS to create different security domains for different groups and tightly control what those tenants can access. Large, shared datasets will be available to authorized OpenStack tenants, just as they are accessible from the HPC cluster.

- **The Outcomes:** At the center of the resource is a large parallel file system that holds around 8 petabytes of configured storage. Surrounding the storage is a cloud of high-performance compute resources with many processing cores and large memory coupled to the storage through an InfiniBand network. Through the use of such technologies as Single Root Input/Output Virtualization (SR-IOV), virtual systems can be provisioned on the compute resources with extremely high-speed network connectivity to the storage and to other virtual systems.

ADAPT's most important feature is the architecture to enable large-scale data analysis through the architectural combination of storage, compute, networking and cloud computing capabilities. The ability to bring the scientist's application to the data and define the environment in which that application runs greatly reduces the friction among the scientist, the data, and the HPC system. This enables scientists to quickly create analysis applications, port them to a very large resource, and have access to extremely large models and observational datasets. Other teams of scientists are using the ADAPT

system and Cycle Computing software on public cloud infrastructure to perform a tree and shrub census and calculate the vegetation biomass and stored carbon within a coast-to-coast swath of sub-Saharan Africa. This biomass estimate will reveal how much carbon is stored in the region's vegetation and could be released as carbon dioxide if those plants burn or die and decompose due to natural or human causes.

This research initiative, which is extremely compute- and data-intensive, was propelled forward with the quick scalability of cloud-based resources. This cloud link, enabled by Cycle Computing's CycleCloud software, gives the NCCS researchers access to the additional compute and storage capacity they need to accelerate their scientific investigation. It also helped land a 2015 HPCwire Readers' Choice Award for "Best HPC Collaboration Between Government and Industry."

Today, scientists interact with the ADAPT team to provision resources and launch ephemeral VMs for "as needed" processing. The data-centric, virtual system approach significantly lowers the barriers and risks to organizations that require on-demand access to HPC solutions. As OpenStack is introduced into the environment, scientists will become more self-sufficient in deploying VMs.

Final Thoughts

Cloud technologies are continuing to impact almost every industry in our business world today. In fact, 74% of tech chief financial officers (CFOs) say cloud computing will have the most measurable impact on their business in 2017, according to the 2017 BDO Technology Outlook Survey.

Keeping up is never easy. However, with good technologies and great partners, you can design and manage cloud and big data analytics, using automation to achieve quality, consistency, and efficiency. Remember, these tools are here to make your life easier by simplifying complex operations and management. Big data will continue to grow, the reliance on cloud will only rise, and today's business will rely even more on IT to support evolving initiatives. Tools like those from EMC Dell, Intel, OpenStack, Red Hat and others can help revolutionize your data center to meet the demands of today's digital economy.