The DevOps Paradigm with Cloud Data Analytics for Green Business Applications

Michael J. Pawlish[1] and Aparna S. Varde[2]
1. School of Business, Adelphi University, Garden City, NY, USA
2. Department of Computer Science, Montclair State University, Montclair, NJ, USA
mpawlish@adelphi.edu, vardea@montclair.edu

ABSTRACT

This paper reviews the emergence of the DevOps (development and operations) paradigm in the industry and the influence it has along with cloud based data management and analytics in the greening of business applications. It considers the geoscience domain as an example discussing usefulness in a GIS (geographic information system). Similar claims can be applied to other domains. Investigating the emergence of DevOps technologies and examining the dramatic shift in IT towards cloud and hybrid models for data analytics, the paper paints a picture of systems that have the ability to green their impact on society. It also addresses concerns from a privacy and security perspective and concludes with open issues for further research.

Keywords

Cloud Computing, Hybrid Models, DevOps, Geoscience, Green Energy, Privacy, Security

1. INTRODUCTION

In the last decade there has been an interesting and profound movement in the information technology (IT) sector that has resulted in the coming together of Development and Operations divisions of software companies [20]. This combined with cloud data analytics can cause rapid growth in IT. The emergence of this converging trend will continue to increase the efficiency and effectiveness of the IT industry, and simultaneously this trend will continue to have repercussions throughout society.

The increased speed of building and testing software on a cloud platform will continue to benefit society across industries and cultures by bringing together market tools for hopeful future solutions to challenges and problems [10]. Some of these benefits are discussed in this paper and the green or environmental aspects are highlighted in the work presented as a survey here.

This survey paper provides a background on the convergence of the Development and Operations divisions known as *DevOps* [1]. It thereby moves to present some of the green energy perspectives that includes cloud and hybrid models with reference to data analytics and IT management. For specific applications, it considers the geoscience domain, explaining the use of DevOps and cloud data analytics in the area of geo-informatics [16]. It considers the context of a geographic information system (GIS) in particular [8]. It also discusses some of the concerns with these advancements and challenges to include privacy and security. In addition, the paper puts forth some open issues with the potential for further research.

This survey paper would be interesting to IT professionals, data scientists as well as business executives. The geoscience community would potentially find it interesting as well. The IT industry would gain an insight into the material presented herewith from a green perspective while also being more acquainted with the focus on DevOps, i.e., merging development and operations. Data scientists would gain insight into the concerned research for defining specific problems based on the open issues and challenges discussed in the paper. Business executives would be able to make better decisions in terms of strategy planning and other such initiatives based on the latest developments and trends so as to maximize the efficiency of their systems and yet maintain the required standards of greenness. Moreover, the security and privacy issues here would interest business decision makers as well. Geoscientists would find all these aspects useful with specific reference to the geo-informatics applications considered herewith.

The rest of this paper is organized as follows. Section 2 discusses the emergence of the DevOps paradigm. Section 3 provides an insight into cloud and hybrid models. Section 4 delves into the green energy perspectives relevant to this survey. Section 5 focuses on the applications of this work in geo-informatics. Section 6 analyzes some of the concerns especially with reference to privacy and security. Section 7 gives the conclusions and outlines potential open issues for research.

2. EMERGENCE OF DEVOPS

The advent of the term DevOps has an interesting history. Until relatively recently the model for developing software and testing it rested in different divisions because most businesses experienced a phenomenon called a *silo effect* [20]. This refers to the prior organizational design where typically development departments were deliberately separated from operations departments for the delegation of work. The outcome was that upper management often had the problem of integrating the two distinct departments. In the development department the employees were focused on designing and writing software while in the operations department the focus was on testing the software for any design flaws. The result of this organizational design was that companies had large divisions with differing

goals that had problems of bureaucracy as well as longer time horizons to finalize the end products. This motivated the need for change. Consequently, management professionals found that by working with smaller teams and blending employees from both operations and development the problems stated herewith were reduced. The result was the appearance of DevOps teams [1, 20]. Given the goal of the end product being a better overall software package developed in a shorter time period, this new organizational design took off in the software industry. The timeline for development was supported by having the DevOps team work together more efficiently and effectively by integrating the former divisions. The overall result was a better end product that was obtained more rapidly. Moreover, when this effort was combined with the advancements in cloud technology it proved to be an interesting enhancement in the software industry.

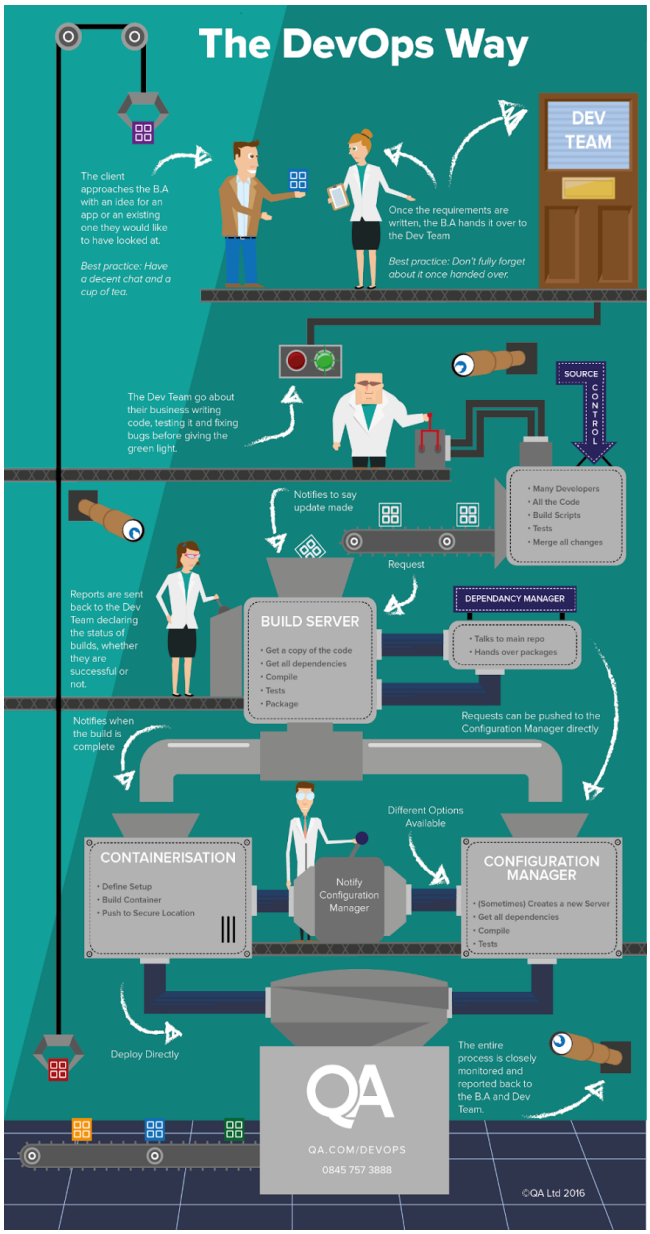

Figure 1: Outline of the DevOps Paradigm in Industry

Presented in Figure 1 is the cultural shift towards DevOps that is underway in the software industry [12]. The shift towards the DevOps paradigm is profound in the sense that the pipeline for software development has been shortened. This has resulted in hopefully a better product being developed in less time by breaking down the tasks into individual components and letting the computers do the more tedious tasks. It has provided a more streamlined approach since development and operations can occur more or less in conjunction with each other, leading to faster and more effective solutions.

3. CLOUD AND HYBRID MODELS

3.1 Cloud Services

The cloud, which utilizes a pay-as-you-go model, offers three main services: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-Service (SaaS). For example, Amazon EC2 is an IaaS, Microsoft Azure is PaaS and Google Docs is a SaaS [10]. In the public cloud deployment model, a feature such as multi-tenancy, that enables a single application to serve multiple customers, is found to be very useful [14].

The public cloud is a good way for businesses to cut IT costs, but still maintain their own *specialized* cloud or datacenter servers to cater to their needs. While the public cloud would not be suitable for sensitive information such as medical records, the private cloud could be a viable option for companies desiring to keep certain data private while also benefitting from some of the many features of the public cloud [10]. However, a private cloud cannot scale elastically like its public counterpart. In that sense, a shortcoming of the private cloud is its limited ability to dynamically provision resources to meet user demand, a feature in which the public cloud certainly excels.

3.2 Hybrid Approach

Given these pros and cons of the cloud models, a hybrid approach which combines some features of both public and private deployment methods can be a viable option for a lot of business processes. Businesses desiring to keep some data on their private cloud or data center servers while also benefitting from the ability of the public cloud to dynamically provision according to the demand would be wise to consider. The concerned business industry can utilize the *on premise* IT infrastructure for sensitive data while non-sensitive data is processed in the public cloud. Therefore, one could consider the hybrid approach as a great way to experience the best of both worlds.

In recent years, the cloud has become a more significant part of our lives, yet despite its rising popularity, studies have shown that many people are unaware that they are using the cloud. Estimates have shown that while 90% of global Internet users have used the cloud although only 29% are actually aware of it [6]. Some popular examples of cloud services are Google Drive, Google Docs, iCloud, Dropbox and Microsoft OneDrive. For example, Google Docs, which started as Google's answer to Microsoft Office, is a widely used example of Software-as-a-Service (SaaS) under the cloud umbrella. The vendor, Google, hosts this application over a network like the Internet. People all over the world can access this application and use it without being concerned about saving the data on their local machine. The data is automatically saved to the cloud servers which can be readily accessed worldwide as long as the users are linked to

a Google account for the data. Thus, from a business perspective, rather than installing and maintaining Microsoft Office on every computer in an office, companies can save resources such as money, HDD (hard disk drive) space and also time by utilizing a SaaS service such as Google docs.

3.3 IT Management and Data Analytics

SaaS applications are just a small aspect of the capabilities of the cloud. Businesses that subscribe to other cloud models can reap greater benefits, especially upstart businesses that cannot handle the initial capital investment of building and maintaining an on premise IT infrastructure. When starting a new business, instead of a company spending an excessive amount of time on backend requirements, such as server management, they can rent the amount of data they need for analysis by subscribing to a cloud data storage service. A common problem with upstart businesses is that the initial investment to build and maintain local servers is too great to handle, and sadly, hinders a potentially great company from attaining its true capacity. Indeed, the initial on premise IT investment alone can hinder an upstart business even before it gets an opportunity to fully begin. By adopting the cloud, an upstart business will have access to a scalable platform to meet their data analytics needs at any given time [10]. As their company grows, and they need more space to house their data, they can simply purchase more. In contrast, on premise IT infrastructures would need to physically build more servers, or find another way to manage their data.

The shift to the cloud has been booming over the past decade since commercial cloud services were offered. To reinforce this growing trend Figure 2 depicts the current market share of cloud services and the year-to-year growth rates for the five largest cloud providers according to the Synergy Research Group [17]. To be noted and presented previously in the paper is the dominance of Amazon with an almost 30% of the total market share of the five largest cloud providers. However, to also be noted are the large year-to-year growth rates, especially at Microsoft (96%) and Google (81%). This growth rate indicates the competitive nature of the cloud industry.

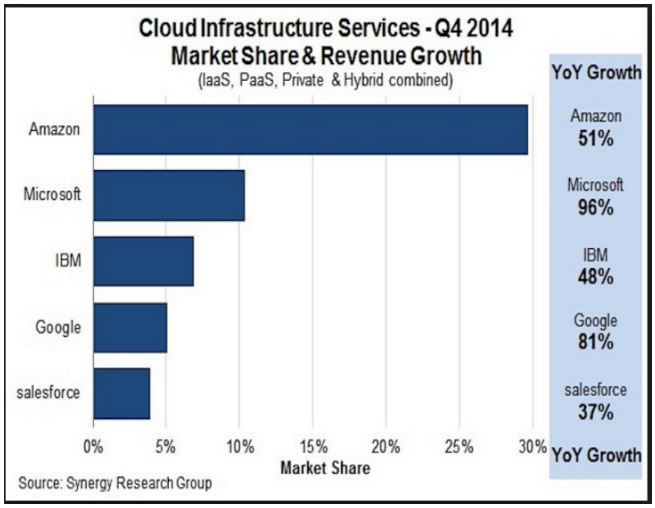

Figure 2: Market Share and Revenue Growth of Cloud Services

According to a study by Gartner, a research and advisory firm providing Information Technology related insight, on premise IT maintenance accounts for around 80% of IT expenditure [10]. As a result, most businesses running and managing their own IT infrastructure spend only 20% of their time on applications and tasks relating to the business. Indeed, the act of maintaining a local IT infrastructure is expensive, not only financially, but also in terms of getting the most out of employees. If this 80-20 rule were shifted in the other direction, the core business needs would be readily met. This is precisely where the adoption of cloud for data storage, processing and analysis can help an organization. Business owners entrusting cloud vendors to house the data of their companies can potentially enable their employees to focus more on core business tasks which will reap greater benefits for the industry.

3.4 Examples of Usage

When adopting the Cloud, businesses can focus on operating expenses, rather than capital expenses, such as building an on premise IT infrastructure. Operating expenses are beneficial to an organization because that gives the business the ability to terminate the costs at will. A simple example of operating expenses in relation to cloud data management would be an online retailer being able to purchase more cloud resources on Black Friday to accommodate for the influx of users. Once they get through the demanding day, they can simply terminate the server capacity they purchased from the cloud to store and analyze their data. In contrast, on premise IT infrastructures would need to build an IT infrastructure to handle the excess demand of Black Friday, which would be a capital expense. The expense of building and maintaining an on premise IT infrastructure could very likely offset the potential benefits of Black Friday revenue. Consequently, adopting the cloud for times such as these would be beneficial for any company looking to maximize its profit and minimize its investment.

Prior to adopting the Cloud, Gregory /Richochet, a fashion retailer in New Zealand, had an IT budget amounting to over $30,000 a year. These expenses covered server management, data processing for each of their stores, remote backup services and IT support for their network and head office. These expenses proved to be overwhelming, so they decided to adopt Vend, an online point-of-sale application, which is run under Rackspace, a cloud hosting provider. This resulted in incredible savings in storing, processing and analyzing data, as well as in other costs. Instead of paying over $30,000 a year, they only paid $350 a month, which amounted to $4200 a year. Consequently, their adoption of the cloud resulted in savings of over $25,000! Indeed, cloud data management can provide substantial savings and be of great benefit to small businesses as well as large enterprises.

Large cloud datacenters are able to provide huge savings to companies because of their ability to purchase in bulk. For instance, the cost of storage at a medium sized datacenter is $2.20 per gigabyte (GB) per month. In contrast, a large datacenter, like the ones cloud providers use, sees huge savings at rates of $0.40 per GB per month [4]. In addition, large cloud data centers can purchase power, network bandwidth, and hardware at 1/5 to 1/7 of the prices offered at medium-sized datacenters, which means that they can give huge savings to consumers and still enjoy a healthy profit [4]. Thus, we can see that the deployment of cloud and hybrid models for the management and analysis of data as well as other IT services distinctly proves beneficial.

4. GREEN ENERGY PERSPECTIVES

4.1 Datacenter Greenness

Over the past decade, global warming has become a concern for many in our society. Since the rise of the industrial revolution, mass production has utilized coal as part of manufacturing goods. This holds true for the computer industry as well. Many of our datacenters, as well as the cloud industry, are run by electrical utilities that rely on coal for power. These electrical generators that are based on coal emit harmful gases that have long-term implications to the environment, as studies have shown. As a result of both future pending legislation in the USA, and strategic factors to have less of an impact on the environment, many businesses, including several cloud vendors, have taken a green initiative. A central question is whether the cloud is less harmful to the environment when compared to using on premise IT solutions. In short, the answer to that question seems to be that the cloud has less of an impact on the environment compared to the scenario of "what if all of the current cloud users were using private data centers" as explained below.

Datacenters consume large amounts of energy, about 1.2% of the electricity consumption in the United States and about half of that energy goes to cooling processors [2]. Since so much energy goes towards cooling the servers, a lot of energy is wasted, and consequently, the datacenters produce carbon emissions that could otherwise have been avoided. To avoid such emissions, the cloud could simply allocate their resources to fewer servers, especially those that are not being used could be switched off to save energy, cooling power and reduce global emissions as found in interesting studies worldwide [11]. Additionally, a 2007 study for Gartner estimated that the Information and Communication Technology (ICT) industry is responsible for 2% of the total global carbon dioxide (CO_2) emissions. As one can imagine, these global CO_2 emissions have become a cause for concern to many environmentalists. In a European Union report, they advocated that a decrease in emission volume of 15-30% is required by 2020. As such, many businesses, including cloud datacenters, need to do their part in decreasing the global emission volume.

While the cloud datacenters need to work towards reducing these emission volumes, "…a research study by Accenture shows that moving business applications to the cloud can reduce the carbon footprint of an organization". According to this report, small businesses reduced their emissions by 90%, while it was found that large businesses reduced their emissions by 30-60%. Indeed, for the businesses passionate in their desire to preserve the environment, it would be wise to consider the cloud.

Although the emission of carbon by cloud datacenters is arousing concern among many environmental scientists, the cloud exhibits features that enable green computing [11]. Multi-tenancy is a key feature in the ability of the cloud to save energy and emit less greenhouse gases in the atmosphere [14]. In essence, multi-tenancy enables *virtualization*, which thereby helps the cloud group computing resources utilized by several businesses. In this way, the services that are used are consolidated to the fewest amounts of physical servers possible, enabling underutilized servers to be shut down temporarily. This means less energy is used from the servers and the cooling systems. As a result, significant reduction in carbon emissions can be achieved.

4.2 Cloud Features for Green Datacenters

There are four key features of the cloud that enable green computing: Dynamic provisioning, Multi-tenancy, Server utilization and Datacenter efficiency [11, 14]. These are briefly explained next.

1. *Dynamic Provisioning:* As mentioned earlier, the capability of the cloud to provision resources based on the amount of traffic it receives is a key feature to energy and cost savings. By utilizing only the server capacity that is necessary, less power and cooling resources are consumed in the datacenter. Thus, lower carbon emissions are released into the atmosphere.

2. *Multi-tenancy:* This refers to multiple companies being able to utilize the software installed on a single infrastructure, such as with SaaS services. This is much more energy efficient than utilizing the software installed on multiple infrastructures.

3. *Server Utilization:* Through adequate use of servers, the applications and resources can be hosted on a single server, which in turn leads to less energy being consumed by the datacenter. This implies less carbon emissions into the atmosphere.

4. *Datacenter Efficiency:* A PUE (Power Usage Effectiveness) level measures the efficiency of energy use. It is measured by the total amount of energy produced by the facility divided by the total amount of energy the IT equipment produces. Cloud datacenters can achieve PUE levels of 1.1 to 1.2, which is about 40% lower than the average individual datacenter.

The cloud has many features that precipitate green computing. While there is certainly room for improvement, such as assuring datacenters only use energy that is necessary, overall, the cloud provides a service many green businesses can use.

4.3 Utilization Analysis of Datacenters

Additional information that is important to note from our research on private data centers is the amount of time the servers are actually being used, in other words, *utilization rate*. We have conducted research pertaining to this on a mid-size university located in the eastern part of the United States over a five-year period [13]. Presented herewith are Table 1 and Table 2 that introduce the concept of utilization rates on two hosts for the first half of a typical year. The year shown here is 2012. The average utilization rate is the amount of time the servers are active.

It is to be noted that the range of the average utilization rate from 29% to 42% translates to the servers not being used more than 50% of the time [13]. This is an important finding since if servers are designed with some type of sleep mode, there is a potential for energy savings. The two main energy costs are for the servers and for cooling. From our research, it has been found that the two main energy costs each account for approximately 50% of the energy costs. Designing servers with a sleep mode when not being utilized would have the dual benefit of lowering

the energy use and also decreasing the carbon footprint of the organization.

Table 1: Utilization Rates for Host 1

2012	Host 1		
Month	*Average Utilization rate*	*Monthly low*	*Monthly high*
Jan.	*38%*	*7%*	*86%*
Feb.	*34%*	*10%*	*85%*
March	*30%*	*7%*	*60%*
April	*35%*	*8%*	*68%*
May	*35%*	*10%*	*63%*
June	*29%*	*9%*	*60%*

Table 2: Utilization Rates for Host 2

2012	Host 2		
Month	*Average Utilization rate*	*Monthly low*	*Monthly high*
Jan.	*42%*	*20%*	*86%*
Feb.	*35%*	*25%*	*90%*
March	*38%*	*21%*	*87%*
April	*35%*	*9%*	*82%*
May	*38%*	*21%*	*84%*
June	*42%*	*18%*	*90%*

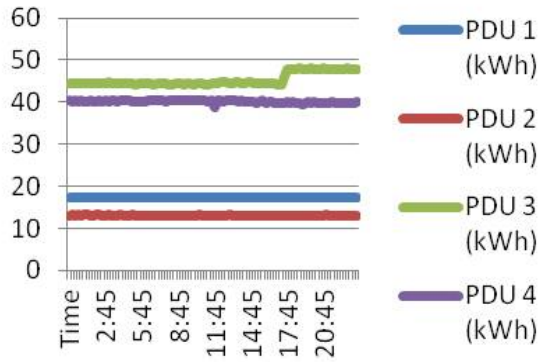

Figure 3: Plots of kWh vs Time for Power Distribution Units

An additional concern that we have noticed in our research on private data centers is the balancing of servers on the Power Distribution Units (PDU) [13]. Presented in Figure 3 are the four PDUs that monitor the power going to the data center servers on an individual day. We have found that the power is consistent from day-to-day, however the power is not evenly balanced on the individual PDUs (as noticed by PDU 3 and 4 using more power per kWh). We believe that by balancing the power on the PDUs in a more even format would prolong the life of the equipment. It would possibly distribute the heat of the individual servers more evenly which would be more efficient.

5. GEO-INFORMATICS APPLICATIONS

5.1 Cloud Data Storage for Geo-Informatics

The area of geo-informatics deals with processing and analyzing information of a geographic nature where spatial and temporal factors are crucial and location-specific services are likely to be needed [16]. It is an important area in the whole realm of geoscience in general. Cloud data management can play a significant role here. Some of the important trends in cloud data services have been presented herewith. Companies that implement cloud technology are using less hardware and licensed software. The reasons behind this shift are the freedom to reach the devices from anywhere, reduction of the investment on assets, lower costs and the ability to shift from a capital expenditure to an operational expense [4, 7].

An overall trend is that more companies will adopt a cloud strategy for storing and analyzing their data in geo-informatics applications and cloud spending will increase. It was foreseen that cloud market would continue to grow in 2016 which has already been witnessed. According to new projections, the global market will grow from its $49 billion that existed in 2016 to $67 billion in 2018. As more organizations shift to some form of cloud services to store and analyze their data, the demand will continue to grow and the amount of computing power will expand [7]. This will have a positive impact on the applications in geo-informatics since they will be able to utilize the increased potential of cloud technologies to enhance their location-specific services.

A second trend is that hybrid models will continue to be a popular option and grow at a rapid pace [13]. Public clouds offer cost savings and convenience, but some companies require the ability to add layers of security and customization, so they opt for a private cloud. Increasingly, purchasers are choosing the best of both worlds: a hybrid cloud that can house non-sensitive data on a public cloud while storing crucial systems and sensitive data on a private cloud or their own datacenter servers. Recently, private and hybrid cloud adoption rates were slightly less than public cloud growth. Moving forward, hybrid options will grow at a faster rate as more companies choose the convenience of the public cloud and the security of the private cloud. Geo-informatics applications will benefit from these hybrid options since they can maintain their critical data for highly sensitive location-specific services on a private cloud and outsource operations of a more generic nature for other less crucial location-based services to the public cloud, thereby incurring energy savings while yet catering to privacy and security issues.

Another major trend is that more staff will be involved in the cloud adoption decision process [7]. One of the most important aspects of cloud based data services is convenience. Armed with only a credit card, business unit leaders can instantly acquire the computing power they need to support expanded operations and scale up to handle major projects. This key aspect of cloud technology has revolutionized the industry. Other issues are that

education and training will be very important because as more companies adopt cloud technology, trained staff will be in demand for growing needs of the organization [7]. This will have an impact on geo-informatics by getting geoscientists and cloud professionals involved in decision-making and strategy planning along with corporate business executives.

5.2 Emergence of DevOps in Geo-Informatics

An important aspect of geo-informatics is GIS, i.e., Geographic Information System. Geo-informatics deals with the area of managing the geographic information in general with respect to collecting the data, extracting the knowledge and provisioning it to users [16]. A GIS deals with the more specific tasks of an IS, i.e., information system per se. It has specific goals with respect to various deliverables such as Demographic Analysis, Emergency Preparedness, Land Management and others [5, 8]. A broad range of these are illustrated in Figure 4 which depicts an overview of a GIS for a town in Michigan [5].

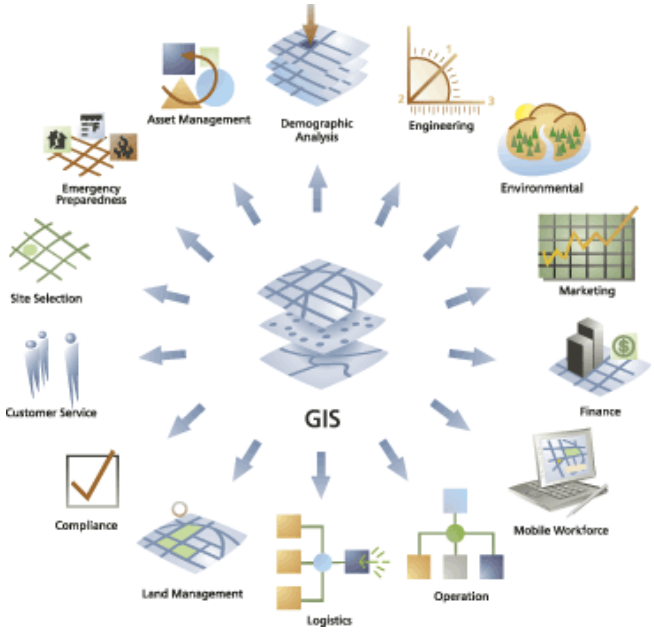

Figure 4: Example of a GIS Snapshot for a town in MI, USA

DevOps, a cultural and professional IT movement focused on changing the mindset of how organizations function, will certainly make a significant impact across many industries and companies in the near future [20]. An example of the broad impact DevOps has is in the Geographic Information System (GIS) area. The ability to portray maps on multiple layers has profound changes in the areas of engineering, demographic analysis and land management to just name a few. The DevOps paradigm provides faster delivery of product features, faster resolution of problems and continuous software delivery, all of which are advantageous in GIS applications. DevOps can benefit the product lifecycle of a company, its competitive advantage and its ability to meet customer needs more rapidly. Despite all of its benefits, DevOps is still in its early stages, because it is about making a significant cultural shift in how an organizations works and the tools the organization uses to develop its products.

A significant trend in the DevOps category is an increase in modular approaches to system building [15]. Previously, IT companies created monolithic products for their customers. However, the best practice today is to arrange small and agile teams to manage individual applications. Instead of huge teams working on one big application, a shift in acceptance in the software industry is occurring where small but agile groups are doing better work in individual applications. This is useful in building systems in geo-informatics where location-specific Development and Operations tasks can be carried out by smaller teams in those respective locations, thereby enhancing product delivery and functionality.

DevOps enable further advances in programmable infrastructure as a code. The idea of automation per se is not novel, however the ability to provision infrastructure easily and seamlessly has changed the nature of the game. As DevOps and the open-source software movement becomes more readily accepted, infrastructure can be programmed so that teams can develop the software and operate in the external environment simultaneously [20]. This is a huge advantage in geo-informatics where interacting with the external environment is essential and frequent.

An additional trend is that developers will take greater ownership of the entire product lifecycle. As DevOps-ready tools see more adoption and out-of-the-box functionality, the traditional silos between developers and operations will lose their importance. As many teams focus on continuous delivery and improvement, it means greater accountability and ownership from developer teams to build and run their solutions. DevOps leaders need to keep developers involved in the performance of the application and any issues that occur, as well as communicate that their job does not end after the application has been delivered [20]. This is beneficial to the designers of GIS since there are several aspects that need to be addressed with development and operations going hand in hand as evident from Figure 3. For example, in order to cater to a goal such as Demographic Analysis, it is important to develop the piece of software that captures the demographics of the region and then conduct the analysis with respect to specific questions that may be of interest to prospective users. Thus, DevOps teams with development and operations occurring in conjunction with each other are useful in this context. Developers need to be part of the entire lifecycle, and have complete visibility into its progress.

On the whole, DevOps will dramatically reduce the time to deploy any feature to production [15]. This trend is already occurring, as product methodologies transition from the traditional waterfall to more agile ways of working. As this transformation occurs, so also will the systems be more risk tolerant. Any changes are less likely to negatively impact the entire system, thus the time to address them will be reduced to hours and minutes, rather than days or weeks. Companies should start measuring how long it takes to deliver a bug fix or a feature to production, and monitor this on a weekly basis. This is a clear metric that can help monitor the success of DevOps [15]. It is particularly useful in geo-informatics applications where location-specific services could be subject to several changes as things progress. Thus, implementing those changes should be a fast process. Analysis of data in GIS and other geo-informatics applications needs to take into account rapid changes in order to reveal useful and timely results. For instance, consider Emergency Preparedness with reference to Figure 4. In order to

develop and operate the system with reference to this goal, it is crucial to incorporate changes that occur in emergency procedures in specific locations. Else, the system may not function well if it is not aware of novel features available, e.g., a new kind of evacuation procedure. Thus a good GIS should be able to incorporate significant changes during the developmental and operational phases. Hence, the emergence of DevOps is highly suitable to speed up processes here in order to meet the goals of the systems and deliver better outputs.

6. PRIVACY AND SECURITY

6.1 Significant Concerns

While the cloud enables scalability, reduces cost and has positive effects on the environment, one of the obstacles preventing businesses from adopting the cloud, is security. While the cloud has been gaining popularity in recent years, there is still some apprehension on adopting the service, for the fear that the data is not secure on any of the cloud deployment models, especially the public cloud [9]. The traditional security methods, such as basic user authentication and authorization, are no longer adequate. This is further aggravated when cloud data management is used in conjunction with DevOps. Since development and operations teams work together to deliver the outputs, the privacy and security issues would be spread across both these teams. We discuss these with specific reference to the SaaS, PaaS and IaaS services provided by the cloud that would be used by the DevOps teams in various applications.

SaaS gives consumers the ability to run the provider applications over a cloud infrastructure. With this model, the security lies with the cloud provider, not with the user. With respect to PaaS, developers do not have access to the underlying layers, thus the burden or security rests with the provider rather than the user. In contrast, IaaS, which serves as a foundation layer for the other delivery models, the security burden rests with the cloud user. While the providers still manage virtualization, servers, HDDs, networking and storage, IaaS users are responsible for their own security, operating system, applications and middleware.

One of the concerns with the cloud is the dependency that each cloud computing model has on each other. For instance, since PaaS and SaaS are hosted on top of IaaS, any security issue that occurs on the IaaS level could have dire consequences on the other levels as well. This is also true the other way around. For example, PaaS offers a platform to host SaaS services, thus the integrity of the PaaS platform is a factor in the SaaS security. A security concern of SaaS lies in Web applications and the manner in which its flaws can harm the SaaS layer or its breaches in security can occur through Web applications. Furthermore, while multi-tenancy is a key feature in green computing and cost reduction, there are security concerns. A key feature of multi-tenancy is the ability for multiple applications to be stored on the same physical server or database. This helps reduce cost and energy. However, it also comes with a security risk. "Since data from multiple tenants is likely to be stored in the same database, the risk of data leakage between tenants is high" [9]. This poses significant issues.

PaaS delivers applications over the Internet. It is a platform where developers can build and customize applications. An example of PaaS is Apprenda, which is a free PaaS service that provides developers with the tools to create mobile and cloud-based applications before using the application in their own on premise infrastructure [3]. While PaaS is a powerful and useful way to create applications, developers need to be mindful of the changes within the platform. Any changes to the PaaS components can compromise the security of the applications. This should be addressed especially taking into account the DevOps paradigm given that development and operations teams could be availing of multiple platforms over the cloud using PaaS while performing their respective tasks.

IaaS provides virtualized computer resources such as servers, storage and networks. In IaaS, cloud users have greater control of their security, however, the Cloud provider controls the network and storage infrastructure. One aspect is that the use of virtualization is a security concern, since there are more points of entry for attacks than physical servers. In addition, since virtual machines on the same server share CPU, memory etc., if there is any malicious machine located on the same server, it could retrieve sensitive information. This is of great concern while using infrastructure as a service. DevOps teams need to take into account virtualization and related aspects while they function together in the industry. Security measures need to be proposed to deal with such issues.

6.2 Research Perspectives

Researchers and developers have often investigated the privacy and security concerns associated with the use of cloud technologies and have discussed potential solutions in the respective applications. For example, Tancer et al. [18] have presented some useful insights on these with respect to the use of the Medical Markup Language, MML, on the cloud. They consider issues such as: consistent availability of cloud systems when tasks are crucial; storage of private medical records on the public cloud; and differences of laws in countries across which medical records would be accessed for practice and research. Some of the arguments applied to cloud storage and processing of the crucial medial data therein [18] could also pertain to sensitive information in other fields such as geoscience. Accordingly, users of many systems such as a GIS need to be aware of such concerns while deploying cloud services, public and / or private, in their systems. Often these concerns can be addressed and solved by the respective developing organizations before providing cloud data services to the users. The responsibility therefore should ideally be assigned to the providers as opposed to users regarding the cloud services and their associated concerns.

Doctoral students in data management and related areas such as data mining find it interesting to conduct research on issues pertaining to cloud technologies. An overview of such dissertation research is presented in [19]. It is found here that many security and privacy issues arouse interests among doctoral candidates. They address these with respect to solutions such as modeling trust in a cloud context; conducting data perturbation to conceal identity while yet maintaining authenticity of results; and incorporating domain-specific aspects from a user perspective. Much of this work presents the potential for future research as well. We will discuss this more in the next section.

The DevOps paradigm being relatively new does not seem to be very extensively researched with respect to privacy and security. However, we advocate that particularly when being used along with cloud data management, the privacy and security concerns can be further heightened due to multiple professionals from different spheres working together and also due to the use of

shared software services, multiple platforms and infrastructure issues. This mandates further research in the areas of cloud data analytics along with DevOps to cater to modern business needs.

As with every technological advancement, there are pros and cons. While the cloud offers several great features for cost reduction and energy efficiency, there are security concerns within each cloud computing model that prospective users should know. Outlining some of the security issues is not an attempt to deter anyone from adopting the cloud, but rather, to arm them with the knowledge necessary to make an informed decision. If a business does decide to adopt the cloud, they should be aware of all the potential pros and cons and thereby outline their strategies. Moreover, if the business is heading towards the adaptation of DevOps in their functioning, they need to understand that some of the privacy and security concerns become even more pronounced. Hence, they need to make business decisions accordingly.

It is important to note that DevOps provides faster and more effective solutions and end products on the whole [15, 20]; cloud data management is beneficial from a green energy perspective [10, 11, 13]; using the two in conjunction would imply balancing productivity very well with greenness; yet privacy and security issues are very prominent [18, 19] especially while using these technologies together. This motivates the need for further research on several grounds, including privacy and security issues, greener solutions and the enhanced usage of DevOps along with cloud services for data management and analytics.

7. CONCLUSIONS AND OPEN ISSUES

The merging of DevOps and cloud technology has brought forth great strides in the software industry that has strong repercussions throughout all industry and modern society. The switch towards cloud usage for storing, processing and analyzing data has increased a trend towards eco-efficiency in datacenter operations that has resulted in less energy being used.

The trend towards more eco-efficient cloud providers is documented by generally lower PUE (Power Usage Effectiveness) ratios in cloud providers than private datacenters. In addition, the organizational hangs in both the move towards the cloud and the merging of the development and operations departments into DevOps has brought forth faster response times to the building of software. Increasingly, the advancement of software is becoming more important in the daily lives of individuals in modern society, and all types of organizations.

Considering DevOps, the datacenter industry and cloud services there are a number of open issues for research as listed below:

- Investigate specific DevOps case studies on where the development and operations departments have been merged to form more effective organizations; and contrast these with the traditional methods
- Address current developments in the cloud market with the leading five cloud companies presented herewith.
- Find solutions to each of the security issues pertaining to the IaaS, SaaS and PaaS services along with DevOps.
- Incorporate heightened privacy concerns especially for location-specific applications such as GIS.
- Apply more enhanced virtualization procedures in organizations that head towards gaining eco-efficiency and reducing the carbon footprint of operations in various datacenters.
- Perform a detailed comparison of several datacenters incorporating DevOps with cloud and hybrid models.
- Study the differing ways in which a datacenter can go into sleep mode by powering down the servers that would be of interest to save energy and investigate these with respect to the DevOps paradigm.

Thus, while the advancements documented in this survey paper are central to IT management and data analytics, there are still open issues that need to be addressed. The security of the cloud will continue to be a concern that must be investigated, especially given the emergence of the DevOps paradigm and some sensitive applications, e.g., location-specific GIS services. The rules and regulations, as well as their enforcement will be an issue that society will need to monitor and enforce into the future. It is expected that this survey paper will be useful to data scientists, IT professionals, business executives and potentially geoscientists as well. The open issues presented herewith would provide the opportunities for joint research in these areas that would further enhance various technologies in DevOps, green IT, geo-informatics and cloud data analytics.

8. ACKNOWLEDGMENTS

The authors thank Cem Davutoglu and Joseph Grillo from the Department of Computer Science at Montclair State University for their inputs in some parts of this survey. The authors' research in related areas has been partly supported by a grant from PSE&G (Public Services Electric and Gas Company) in NJ, USA.

9. REFERENCES

[1] Agile Admin, What is DevOps, theagileadmin.com/

[2] Ahmad, I., Ranka, S. *Handbook of Energy-Aware and Green Computing*. Boca Raton, FL: CRC Press (2012), pp. 899.

[3] Apprenda, https://en.wikipedia.org/wiki/Apprenda

[4] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I., Zaharia, M. *Above the Clouds: A Berkeley View of Cloud Computing,* Tech. Rep., Electrical Engineering and Computer Sciences, UC Berkeley (2009).

[5] Dangermond, J. *Muskegon County GIS*, ESRI, MI, USA, http://www.muskegoncountygis.com/about.html (1998-2016).

[6] Danova, T., *Most People Are Still Confused about Cloud Storage, And No One Is Winning the Race to Educate and Acquire Users*, Business Insider (Aug 2014).

[7] De Bruin, B., Floridi, L. The Ethics of Cloud Computing, *Science and Engineering Ethics* (2016), Volume 23, Issue 1, pp. 21-29.

[8] Goodchild, M.F. Geographic Information System, *Encyclopedia of Database Systems*, Springer (2009), pp.1231-1236.

[9] Hashizume, K., Rosado, D.G., Fernandez-Medina, E., and Fernandez, E.B. An Analysis of Security issues for Cloud Computing, *Journal of Internet Services and Applications,* Springer London (2013), Volume 4, Issue 5, Print ISSN 1867-4828, Online ISSN 1869-0238.

[10] Kepes, B. *Cloudonomics – The Economics of Cloud Computing* Tech. Rep., Diversity Limited, Rackspace Hosting (2011).

[11] Kumar Garg, S., and Buyya, R. *Green Cloud Computing and Environmental Sustainability,* Tech. Rep., Cloud computing and Distributed Systems (CLOUDS) Laboratory, Department of Computer Science and Software Engineering, University of Melbourne, Australia (2011).

[12] Mclvor, K. *The DevOps Pipeline, QA Limited,* Tech. Rep., Berkshire, England (Feb 2016).

[13] Pawlish, M., Varde, A.S., Robila, S.A, and Ranganathan, A. A Call for Energy Efficiency in Data Centers, *ACM SIGMOD Record* (2014), Volume 43, Issue 1, pp. 45-51.

[14] Rouse, M. What is multi-tenancy? – Whatis.com, Aug 2014.

[15] Roche, J. Adopting DevOps Practices in Quality Assurance, Communications of the ACM *(2013)*, Volume 56, Issue 11, pp. 38-43.

[16] Sinha, A.K. *Geoinformatics: Data to Knowledge* (2003), Geological Society of America.

[17] Synergy Research Group, Strategic Market Intelligence. Reno, NV, https://www.srgresearch.com/

[18] Tancer, J. and Varde, A. The Deployment of MML for Data Analytics over the Cloud, *IEEE ICDM Workshop*s, Dec 2011, 188-195.

[19] Varde A. and Tatti, N. A Panorama of Imminent Doctoral Research in Data Mining", *ACM SIGMOD Record* (2014), Volume 43, Issue 3, pp. 71-74.

[20] Walls. M. *Building a DevOps Culture*, (2013) O'Reilly Media.

About the authors:

Michael Pawlish is currently an adjunct professor at Adelphi University in the Robert B. Willunstad School of Business located in Garden City, NY. He obtained his PhD in Environmental Management from Montclair State University, NJ, USA and his MBA from San Francisco State University, CA, USA. His research interests are business-sustainability focused operations including eco-efficient datacenter processes and sustainable entrepreneurship. He has publications in these areas spanning multidisciplinary venues encompassing Computer Science, Environmental Management and Business. His PhD dissertation has been supported by a research grant from PSE&G in NJ.

Aparna Varde is a tenured Associate Professor of Computer Science at Montclair State University, NJ, USA; and Doctoral Faculty in their Environmental Management PhD program. She obtained her PhD and MS from Worcester Polytechnic Institute, MA, USA and her BE in Computer Engineering from University of Bombay, India. Her research interests span Data Mining, Artificial Intelligence and Database Systems with emphasis on multidisciplinary work. Her current research projects entail commonsense knowledge, smart cities and cloud data mining. She has around 80 publications in journals such as AIEDAM, MTAP, TKDD, IJCAC, SIGMOD Record, SIGKDD Explorations; and conferences/workshops of AAAI, SIGMOD, CIKM, KDD, SDM, ICDM, ICDE, EDBT etc. She has served as a panelist for NSF in their division of Intelligent and Information Systems; a reviewer for journals (TKDD, TKDE, DMKD, VLDBJ, DKE, AIEDAM etc.) and PC member of conferences (EDBT, ICDM, CIKM, KDD etc.). Her research is funded by grants from organizations such as NSF, PSE&G and other private companies.