Addressing Command & Control (C2) Complexity of Multicloud Infrastructure

ATARC Multicloud Working Group

December 2022

Advanced Technology Academic Research Center

Acknowledgements

On behalf of the Advanced Technology Academic Research Center, I am proud to announce the release of the 2022 Report titled **“Addressing Command & Control (C2) Complexity of Multicloud Infrastructure”**, authored by the members of the **ATARC Multicloud Working Group**.

I would like to take this opportunity to recognize the following individuals for their contributions:

Kapil Bareja, MIT; Working Group Advisor Chair

Erik Johnson, Cloud Security Alliance

Geoffrey Mershon, UnifyPoint

Jeremiah Sanders, VMWare; Working Group Industry Chair

Manjit Singh, Agilious; Working Group Industry Vice Chair

Sincerely,

Tom Suder, Founder, Advanced Technology Academic Research Center (ATARC)

Disclaimer: *This report was prepared by the ATARC Multicloud Working Group members in their personal capacity. The opinions expressed do not reflect any specific individual nor any organization or agency they are affiliated.*

Table of Contents

1 Multicloud Complexity

Today's multicloud environments are highly complex and consist of infrastructure spread across owned data centers, colocation data centers, edge deployments, and/or leased from public cloud providers. This infrastructure is made up of millions of components and software that are responsible for providing reliability and performance for the applications they are running. Warfighters depend on these components to meet required reliability and performance needs for applications. Managing this complexity from Day 2 onward is a huge challenge that is human resource intensive, highly reactive, and fraught with pitfalls.

The life blood of IT encompasses the millions of components, and software that runs on them, that are embedded throughout the infrastructure ecosystem. These components exist within machines and are responsible for application and mission success. Every component in the infrastructure plays a critical role. For example, an HDD failure will decrease the amount of useful data stored in memory which limits the amount of data that can be processed by the CPU and transmitted out the machines network ports. This failure has multiple impacts. First, it directly compromises the performance and reliability of applications, and secondly, it lowers asset efficiencies. These impacts are amplified as component failures, or software that runs on them, occur frequently across fleets of machines that play a major role in mission success. A healthy component fleet delivers high performing and reliable applications to meet warfighter needs, while also maximizing efficiencies across the fleet of infrastructure assets.

Knowledge about fleets of components/machines is critical, however this knowledge is lacking across both commercial and public markets. Understanding the full picture of the infrastructure ecosystem is imperative, including the component type, location, software versions running (BIOS, firmware (FW), operating system (OS), and kernel versions), and the overall reliability of the components, machines and software. This requires accurate and timely information that is auto-updated without human involvement. Instant observability and insight tools are the basics for gaining command and control over growing infrastructure needs. Possessing this knowledge empowers IT workforce to be strategic and enables quick analysis of what is and is not working so decisions can be made quickly for preventative maintenance and further improve mission success.

Gaining command & control for Day 2 operations across multicloud environments is challenging and an area that both commercial and public markets struggle with. The common band-aid solution has been to over-provision humans and infrastructure. Unfortunately, this fix has only amplified the problem. Additional challenges include continuous operations with proactive and predictive maintenance, minimizing cyber-risk, solving IT workforce issues including empowerment and industry shortage of talent, and maximizing operational and asset efficiencies.

2 Reactive IT

Awareness of failures usually arises from end-users of applications. The ensuing response is to take reactive measures after the failure has already occurred. Today's reactive nature to complex environments relies heavily on humans. It is challenging enough to solve failures when applications are on dedicated machines, however today's applications are distributed across tens, hundreds, or thousands of machines running in multicloud and/or edge infrastructure which amplifies the complexity of identifying and resolving issues. Furthermore, today's modern applications are designed to withstand a certain amount of infrastructure failures. This essentially cloaks issues and hides them from IT operations or DevOps teams. As it is impossible to identify and fix all failures that exist, triage and repair teams focus on fixing just enough failures so that the application will continue to minimally work. Unfortunately, this means that the system is likely only a few infrastructure failures away from the issue repeating itself, resulting in another failure.

Administrators rely on a combination of traditional monitoring tools and trial and error when applications fail. These traditional monitoring tools force administrators to sift through mountains of individual machine data with the hope that it points them to machines that "might" be experiencing an issue. Commercial Off-The Shelf (COTS) tools that enable proactive and predictive maintenance are long sought after but do not exist on the open market. Therefore, the only option is to have humans perform trial-and-error triaging that requires them to log into each machine to manually look for something that might be broken. This effort is ripe with pitfalls, which includes the enormous opportunity for human error. A Unify Point, Inc. advisor, who was a lead in Google's data center health, automation and operations tools team, said the #1 failure in Google's infrastructure ecosystem is humans, and those human-induced errors are far more complicated and challenging to resolve than any hardware induced failure.

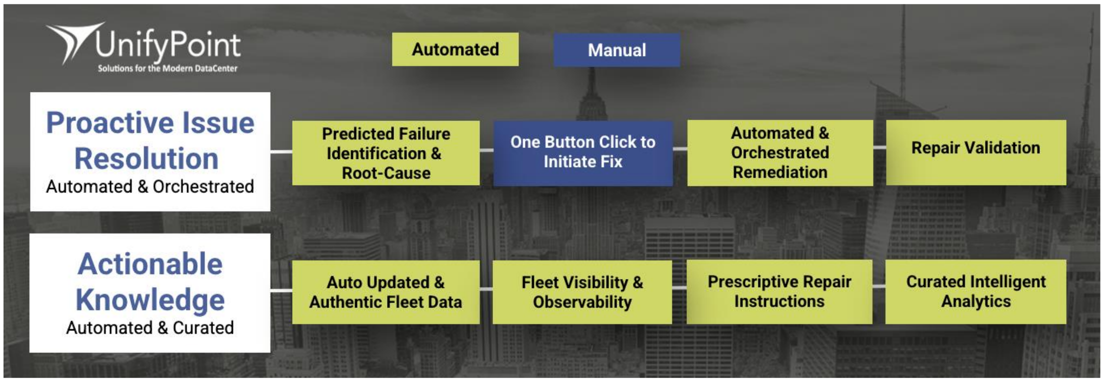
3 Proactive IT

A key to being proactive vs. reactive is the ability to fix issues before they occur, a practice known as preventative & predictive maintenance that includes capabilities like auto-remediation and rich data analytics/knowledge. This includes many different facets of an IT organization, including broad IT teams of operations, incident response, procurement, application, cyber security, SRE and DevOps experts.

Insights from an Arete Research article on IT outlook for 2023 show that budget cuts are inevitable. To combat budget cuts Arete Research identified an area of focus for infrastructure-related items is to invest in transformational initiatives. A growing area of transformational investment is with AI tools. These tools can hyper-automate daily tasks with intelligent software that maximizes efficiencies and can deliver significant budget and resource savings.

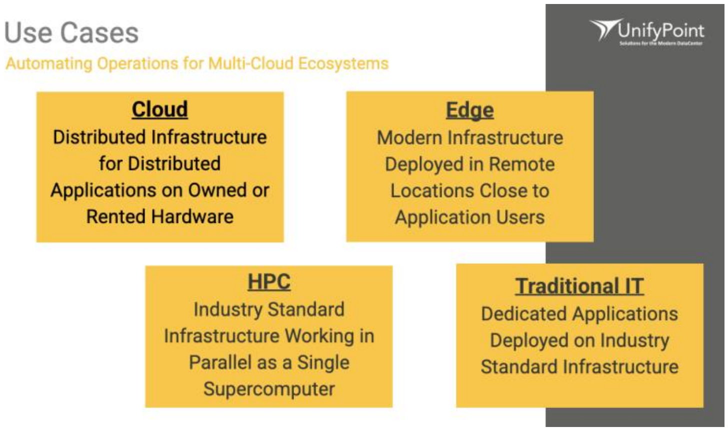
From an infrastructure perspective, the ability to predict hardware failures enables IT operations to resolve issues without impeding applications and missions. Once a failure is predicted, adding the ability to create prescriptive remediation fixes (tickets with detailed repair instructions), along with auto-repair validation, would minimize both repair time and human failures, while allowing use of low to moderately skilled workforce to perform mission critical IT roles. This would free up skilled IT workforce to perform higher value strategic initiatives to better serve customers.

Preventative maintenance is another AI tool needed for transformational investments. Similar to how a medical healthcare worker maintains accurate records of patient health, or auto mechanics maintain health records of a vehicle, it is equally critical to maintain accurate health records of IT infrastructure. This information is often stored in the heads of administrators and/or manually typed into a ticketing system that holds tickets for all issues throughout an organization making it nearly impossible to use for data analytics. Having accurate and intelligent curated analytics information can help a company quickly identify and root-cause flapping or interconnectivity issues that would otherwise go unnoticed, causing erratic application and network behavior. This knowledge enables IT workers to easily compare similar issues and quickly learn how they were previously solved, potentially saving significant time and minimizing the introduction of human failures; some of the most difficult issues to solve. For example, machines can experience repetitive issues where a component appears to be dead, is replaced, and the replacement component works for some time then is deemed failed, and is replaced, and the issue repeats over and over. Without having analytics that point to a bad slot, administrators will continue wasting time and money on something that is not fixable.


UnifyPoint
Solutions for the Modern DataCenter
Automated
Manual
Proactive Issue Resolution
Automated & Orchestrated
Predicted Failure Identification & Root-Cause
One Button Click to Initiate Fix
Automated & Orchestrated Remediation
Repair Validation
Actionable Knowledge
Automated & Curated
Auto Updated & Authentic Fleet Data
Fleet Visibility & Observability
Prescriptive Repair Instructions
Curated Intelligent Analytics

Another example of gaining command and control knowledge for preventative maintenance is to have detailed fleet knowledge for intelligent procurement. Understanding fleet-wide reliability analysis and metrics on systems and components allows IT procurement teams to buy infrastructure that is proven to be highly reliable and high performing. This lessens future failures, performance loss, and impact to Service Level Agreements (SLAs). Insight into the reliability metrics of component fleet is critical as it is highly unlikely that the same component vendors have the same Annual Failure Rate (AFR). If it is known that one of the DIMM vendor models is failing at twice the rate of other DIMM vendors then proactive measures can be taken to discontinue procuring machines with that model in it. For components containing firmware (FW, HDDs, SSDs, NICs, etc.), your intelligent AI tools might show that FW version "A" had an 8% AFR while FW version "B" had a 2% AFR. Armed with this knowledge, two preventative maintenance actions can be taken… 1) component fleet needs to be upgraded from the "A" FW to the proven higher reliable "B" version and 2) procurement teams become empowered to leverage data analytics regarding their environments to implement changes with suppliers to prevent failures and improve application uptime. Actions from intelligent fleet analytics will improve preventative maintenance and provide additional cost savings through operational and asset efficiency gains.

Understanding the health of infrastructure is critical whether it is owned, or rented from cloud vendors. Applications rely on infrastructure for performance and reliability regardless of hardware location or owner. A challenge with renting infrastructure from the cloud is in not knowing the health of the machines. Cloud vendors don't want this known since they are in the business of renting machines trying to maximize occupancy rates and profits. Over time, as infrastructure failures occur and pile up unknowingly, systems may reach a tipping point where applications are impacted. Due to the lack of infrastructure health knowledge, a common industry resolution is to rent additional infrastructure and distribute applications across these machines so applications can meet end-user demands. This is known as "application sprawl" which essentially increases machine over-provisioning. Application sprawl works; however, it increases operational costs and complexity. Having a detailed understanding of machine health and failures (including predictive failures) of rented infrastructure allows proactive measures to be taken immediately without wasting countless resources looking into the myriad of other potential failures that could have caused the outage.


Use Cases

Automating Operations for Multi-Cloud Ecosystems

Cloud
Distributed Infrastructure for Distributed Applications on Owned or Rented Hardware

Edge
Modern Infrastructure Deployed in Remote Locations Close to Application Users

HPC
Industry Standard Infrastructure Working in Parallel as a Single Supercomputer

Traditional IT
Dedicated Applications Deployed on Industry Standard Infrastructure

4 Rich Data Analytics

Investing in intelligent and powerful AI tools with rich data analytics is critical to maximizing operational and asset efficiencies, providing continuous operations “sine qua non”, and freeing up valuable workforce resources to combat the talent shortage plaguing the industry. For hyper-automation with rich data analysis capabilities, it is critical to collect the right data and combine that with domain expert knowledge to make decisions that drive action and results. Traditional IT operations tools (aka monitoring tools) have been around for decades and use data collection standards, called IPMI and SNMP, that were introduced in 1989 and ratified to latest versions by 1992. These monitoring tools were designed to collect as much data as possible and present it for end-users to make sense of it as they don’t know what the data means. It’s unfortunate that these tools are unable to solve a problem given the amount of people needed just to make them run. Complicated tools are not the answer. New tools are needed to identify the right data to collect from components and domain experts, who understand what the data means, are needed to make accurate decisions that drive action and knowledge. This is highly complex. For example, storage component vendors all provide common errors that have the same error label and data value, however that same data value means something different for each vendor. The largest HDD vendor, who only uses their own drives in data center machines, doesn’t even know when their product fails. Once a week they send IT personnel down every isle of their many data centers looking for amber lights. Their employees write down information about the failed HDDs then manually creating incident tickets to repair the failures. This only works for hard failures where lights appear on the outside of machines. Unfortunately, the majority of components do not have failure lights, let alone lights for predictive failures or flapping issues where components fail then miraculously reappear and repeat themselves often without anyone knowing.

5 Cyber Security

The industry's largest cyber risk/threat to both public and commercial markets is humans. Through unintentional errors including triaging, basic mistakes, trial and error, fat fingers, forgetfulness, and laziness, humans unknowingly create vulnerabilities or holes for bad actors to penetrate infrastructure simply through triage efforts, changing infrastructure or OS software, or through configuration modifications. Humans could also create these same vulnerabilities with intention and/or malice. Unfortunately, it can be difficult to determine if actions are intentional or unintentional. What is well known is that limiting human interaction with infrastructure will directly minimize cyber risk.

It is imperative for a company to know exactly what infrastructure is available, where it is located, and what versions of software are running on it. Having detailed, accurate, and up-to-date knowledge at your fingertips is critical for identifying and resolving fleet vulnerabilities quickly. For example, if a cyber vulnerability becomes known regarding a given version of software OS, kernel, BIOS or FW version, having immediate and accurate visibility across the fleet of machines would help a company understand and scope the size of risk while also providing details needed to quickly fix those vulnerabilities. Another example is if a version of software is installed on a machine, which was previously known to have a cyber vulnerability, then having a solution that polices those events and notifies a company of them instantaneously can enable proactive action to close vulnerabilities immediately. A preferred and optimal solution would be one that automatically polices and remediates the issue while providing notification for strategic management awareness for the resolved issue. In addition, this would further minimize the need for humans to log into machines.

6 Industry Workforce and Talent Shortage

A key challenge impacting both public and private industries is the continuous struggle for IT resources and talent. The ability to find talent has been a growing issue with far more demand than resources to fill it. In 2021 the Bureau of Labor Statistics showed 1.4 million unfilled IT jobs in the US. After speaking with Lt. General Skinner (Director of Defense Information Systems Agency and Joint Force Headquarters for DoD Information Network), he believes the true number of unfilled IT job requisitions far exceeds this number. Regardless, it is agreed that automation and/or the ability to use non-skilled resources to perform IT roles is a necessity.

It's no secret that the public market has a difficult task of attracting and retaining talent as they are competing with a private market that is willing to offer higher wages, significant benefits, bonuses, stock options and access to new innovative technology. The public sector's inability to compete for skilled workers forces them to adopt other methods to fill open positions. A commonly used practice is to hire un-skilled workers and train them on the job. This is an expensive and time-consuming fix. In the long run it is hard to retain those employees when their newfound training and experience makes them more attractive to the private market.

A more efficient method to solve this would be to invest in new, innovative AI hyper-automation and orchestration tools that self-manage Day 2 onward operational duties. This would free-up significant IT resources and talent across the industry, while enabling opportunities to leverage non-skilled workers to perform critical IT roles and activities.

7 Workforce Empowerment

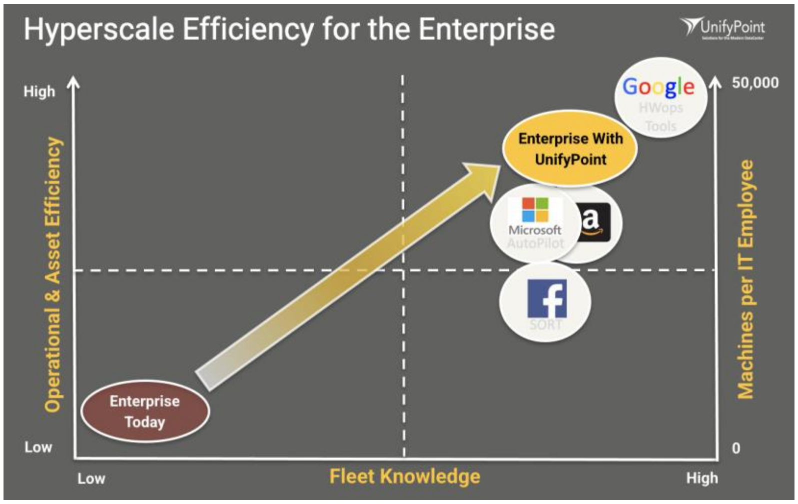
Workforce turnover rates throughout the US are increasing and the IT market is no exception. In 2022, CompTIA estimated there to be 2,010,882 IT employee separations that equates to over 37% annual turnover rate. On top of the turnover rate is the replacement rate, which averages 7% or another 411,622, plus the overall market growth rate, which CompTIA expects to be "twice the rate of overall employment across the economy". As employees depart, they create not only a gap in production, but also a gap in knowledge about the company's IT environment. It can take years for replacement workers to gain the knowledge back that was lost. Maintaining that knowledge in tools would be a more effective and efficient method for storing, maintaining and leveraging that data.

A top concern for many CIOs, according to Arete Research, is trying to reduce IT support burdens. Constant failures in IT environments build pressure for IT workers that can quickly bring down employee morale. The industry shortage of workers has an amplification effect that can increase employee burnout, human errors and demotivate the workforce.

Tools that limit or remove employee stress will minimize turnover rates and improve morale. Moving employees out of the trenches, providing them with rich data for strategic roles and decision making, and enabling them time to work on more strategic initiatives will empower and motivate the workforce and increase retention rates.

8 Maximize Operational Efficiencies

Many opportunities exist to increase operational efficiencies. Identifying the organization's IT efficiency rate is a valuable exercise and easier to undertake than most believe. Through dozens of meetings with commercial corporations and hyper-scaler operations teams, Unify Point, Inc. has identified a simple and straight-forward metric to measure this; dividing the total number of infrastructure machines by the total number of people in an IT organization. IT delivers applications to their customers and these applications require infrastructure to run. The typical commercial market efficiency level was found to be between 20 to 50 machines per IT employee. It is also important to remember that these results are based on commercial and public markets using existing COTS product that includes things like configuration management database, alert monitoring, performance monitoring tools, excel spreadsheets.



Currently, Google has the most efficient IT organization in the world. They have managed to reach efficiency levels of approximately 45,000 machines per IT employee, which is roughly 900-to-2,000 times more efficient than commercial or public markets. This is a staggering difference, and it is important to understand how they achieved these results. Google realized that traditional monitoring tools were both ineffective and inefficient. After finding no available product on the market that enabled them to maximize efficiencies, they invested heavily in a software development team to develop new AI tools to automate and provide them with complete control their growing environments. These intelligent automation tools were found to be highly effective while providing missing intelligence that enabled proactive and preventative maintenance so they could manage their infrastructure with ease.

Recent insights from an Arete Research article on the IT outlook for 2023 show that in today's economy budget cuts are inevitable. A key item to solve for is reducing IT support burdens. New innovative tools are needed for both the commercial and public markets to maximize operational & asset efficiencies. Reaching similar efficiency levels to those of Google will likely be unattainable for quite some time, however increasing efficiencies by even 10 times or more is very realistic and can produce massive savings. With even a 10x improvement, the typical large enterprise environment could save $100 million a year, freeing up significant resources to focus on strategic initiatives that benefit the warfighter. Apply those savings across the entire Federal government and the savings could easily reach over $1 Billion.

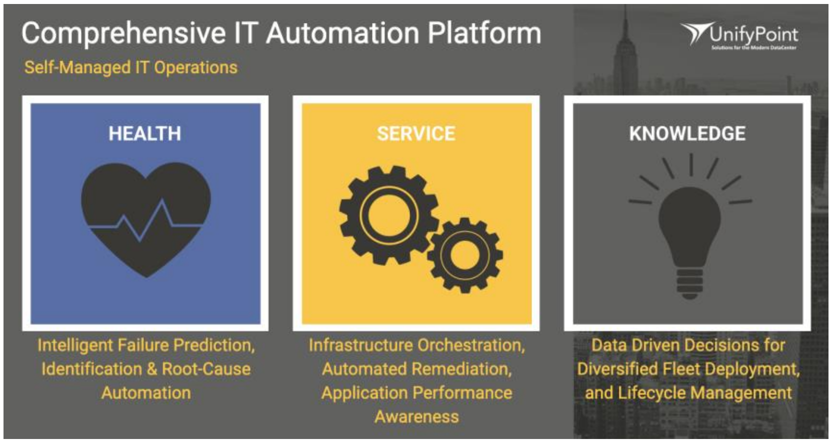
9 Maximizing Asset Efficiencies

Improving asset efficiencies can reduce both CAPEX and OPEX while helping to reach environmental sustainability goals. These benefits apply for both owned or leased (cloud) assets.

For owned assets, the goal is to have a 100% healthy fleet of machines. If every component inside machines is healthy, they will produce 100% of the performance and reliability they were designed for. Having component failures, or imminent component failures, will directly impact the applications being served. For example, a machine has a quad port network card and one of those ports is dead leaving a 25% loss in throughput. These failures can go unnoticed for weeks, months or potentially years. Having a solution that identifies failures real-time and automates remediation with prescriptive repairs across fleets of infrastructure will maximize asset efficiencies.

Understanding of failures in cloud infrastructure is just as critical as owned assets, however the actions to fix them are different given the inability to access to the physical machines. Cloud Service Providers (CSPs) are in the business of leasing machines/infrastructure. The more machines they can rent, the better their revenues and profits are. They are incentivized to have customers rent additional machines if their applications are having performance issues as it increases profits. The last thing they want is for end users to understand that they are paying 100% for machines that are producing less than 100% performance as this creates for an unhappy customer experience and shifts price negotiation leverage away from CSPs, which they don’t like. However, for end users who are renting the machines, it is invaluable to understand the level of health of the product they are buying as it directly impacts their application performance and reliability. This information is important across broad IT groups including system administrators, SREs, DevOps, procurement, and application teams to name a few. For teams who are responsible for applications meeting SLAs, having knowledge regarding health issues in machines they are relying on enables them to take proactive measures to stop leasing the faulty machines and move their applications to new healthy ones. This prevents applications from failing and IT going into default reactive mode to solve issues after application users are impacted and complain. A typical resolution option is the practice of renting incremental machines and distributing applications on the new equipment to meet their SLA metrics. This practice lowers operational and asset efficiencies while at the same time increasing operating costs and complexity. Having this knowledge provides procurement teams with tools to effectively manage their CSPs products and pricing that is tied to results. Afterall, knowledge is power.

10 Solution


Comprehensive IT Automation Platform

Self-Managed IT Operations

UnifyPoint

Solutions for the Modern DataCenter

HEALTH

Intelligent Failure Prediction, Identification & Root-Cause Automation

SERVICE

Infrastructure Orchestration, Automated Remediation, Application Performance Awareness

KNOWLEDGE

Data Driven Decisions for Diversified Fleet Deployment, and Lifecycle Management

Applications are imperative for both commercial and governments to run their businesses and missions. These applications rely on infrastructure to provide the performance and reliability warfighters need. The health of the component fleet equals the health of the infrastructure fleet, which in turn delivers reliability and performance for applications. A new innovative solution needs to self-manage the health, service, and knowledge across fleets of multicloud infrastructure and software, while also integrating into existing service workflow tools to seamlessly blend in without disrupting the flow of operations.

High-level multicloud C2 requirements:

- Maintain a detailed, accurate and updated list of infrastructure fleet hardware and software assets
- Provide detailed search capability across fleets of machines
- Curated and searchable fleet health history
- Ability to automatically identify, predict and root cause infrastructure failures
- Automate and orchestrate failure remediation
- Integration with existing workflow tools
- Automatic creation tickets including prescriptive repair instructions (for owned machines)
- Fix validation and verification (for owned machines)
- Identify and correct human errors in remediation process
- Curated and searchable fleet reliability metrics
- Self-managed capability such that it does not take a team of “experts” to “make it work”
- Supported use cases of on-premise, cloud, co-location and edge deployment environments

11 Solution Benefits

Keeping the IT department lean is essential to maximizing returns on investment, however this has been difficult-to-impossible to do since humans, in many cases, are the default solution. New mission requirements continue to grow, and IT must keep up with them while also continuing to support existing deployments and their growth. This means increasing both OPEX and CAPEX budget demands to maintain and grow service levels. Budgets are rarely able keep up with these demands. Adding even more complexity is the constant battle of finding and retaining workforce in a market where there exists a massive shortage of both workers and talent. This challenge is amplified when budgets fail to match market demands needed to attract talent. The need for automation is more critical than ever, but automation alone is only the first step. Additional items are needed to greatly expand the benefits of automation. Orchestration, along with intelligent analytical knowledge and awareness, are critical to maximizing efficiencies and achieving C2 of growing and complicated multicloud environments.

Orchestration acts as an extra layer of automation that can yield substantial benefits. For owned machines, imagine systems that have failures or predicted failures. Automating the identification and root-cause of these errors across a fleet of machines is a huge benefit, however orchestrating fixes for those issues amplifies operational and asset efficiencies by shortening the time to repair, identifying and resolving human failures in real-time, validating repairs, and enabling use of low to non-skilled workers to perform mission critical roles.

For rented/leased machines, having the knowledge of failures that are impacting your applications is invaluable. This allows for strategic planning to proactively move applications to healthy machines and lower cost and complexity by releasing faulty machines back to cloud vendors. Understanding the reliability and quality of product from cloud vendors is also important. Rich data analytics on reliability/quality issues provides insights for making quick and accurate preventative maintenance decisions. It also aids in vendor management including product/service assessments and comparisons. Intelligent data drives intelligent decisions.

Today’s IT environments are ever-changing and more complicated than ever before. Having curated and accurate knowledge is necessary to maintain control. Data, for the sake of collecting data, is of no benefit or value, but having tools that understand what the data means enables new capability for managing infrastructure strategically and proactively. Intelligent analytics at your fingertips is powerful and empowers personnel to take proactive measures.

Having the benefits of preventative and predicted health maintenance, orchestration for fast remediation of issues, along with C2 knowledge for strategic management will enable IT organizations to maximize their operational and asset efficiencies. Some high-level benefits of such a solution include:

(1) Up to 10X operational and asset efficiencies.

(2) Curated knowledge of your environment with accurate and intelligent data at your fingertips.

(3) Lower application failures.

(4) Higher application performance.

(5) 75% reduction in fleet support labor cost achieved through automation of fleet health and pre-

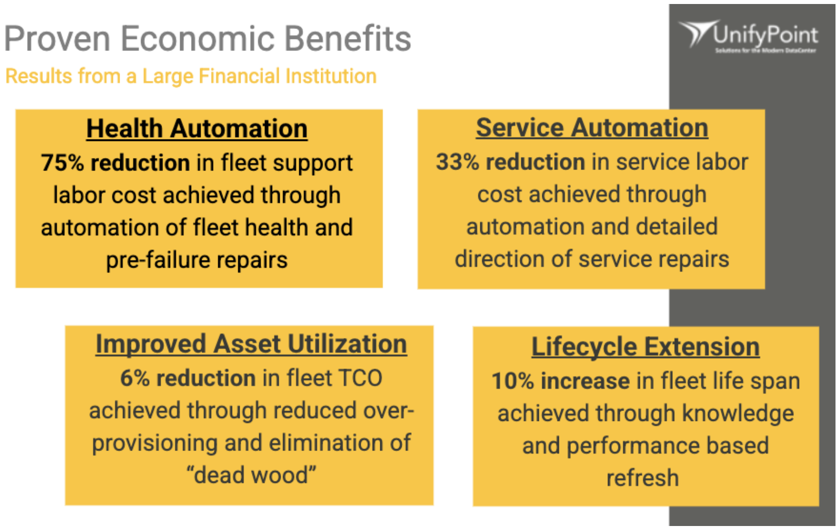
failure repairs.

(6) 33% reduction in service labor cost achieved through automation and orchestration of service repairs.

(7) Reduction in fleet TCO achieved through reduced over-provisioning.

(8) Increase in fleet life span achieved through knowledge and performance based refresh.

(9) Meet sustainability goals by reducing infrastructure needs, power savings and lowering e-waste/refuse.



Proven Economic Benefits

Results from a Large Financial Institution

UnifyPoint

Health Automation

75% reduction in fleet support labor cost achieved through automation of fleet health and pre-failure repairs

Service Automation

33% reduction in service labor cost achieved through automation and detailed direction of service repairs

Improved Asset Utilization

6% reduction in fleet TCO achieved through reduced over-provisioning and elimination of "dead wood"

Lifecycle Extension

10% increase in fleet life span achieved through knowledge and performance based refresh