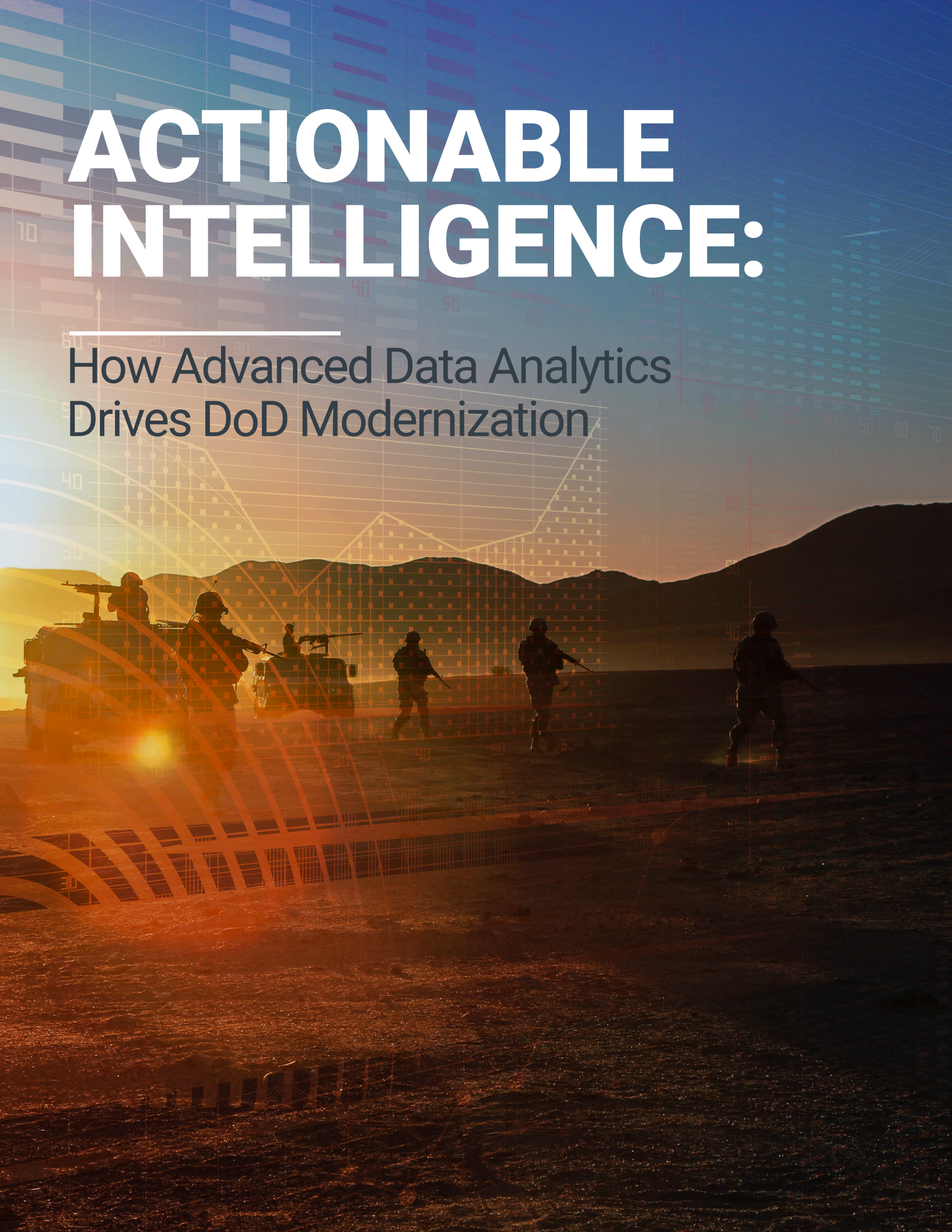
ACTIONABLE INTELLIGENCE:

How Advanced Data Analytics Drives DoD Modernization



INTRO

"Actionable intelligence" has always been one of the military's most important tools, whether in alerting the British to the impending threat of the Spanish Armada or in helping Eisenhower determine the timing of D-Day. And never has actionable intelligence been a more important goal of the Defense Department than it is today.

The U.S. military already collects more data than it can turn quickly into actionable information, but the challenge goes much deeper than military operations. In nearly every corner of DoD – an organization that spans the globe with more than 5,000 locations or outposts, more than 2 million combined active-duty and civilian personnel (and more than 2 million retirees and their family members), and a budget of $692 billion for 2018 – the need to collect, store, analyze and share data is a growing dilemma.

One obvious example: the millions of hours of full-motion video collected every day from round-the-clock drone flights and other sources. The volume of video already outstrips the ability of analysts to parse it, and it's projected to grow by a million times in the next five years, according to the National Geospatial Intelligence Agency, which wants to automate 75 percent of the analysis. Less obvious would be sensors attached to a Humvee, delivering performance data via the Internet of Things to predict necessary maintenance. And in between are cybersecurity monitoring and analytics, procurement and logistics systems, health care wearables and monitors, energy-use sensors, large-scale records systems and a lot more.

Many of the top challenges facing DoD fall into that area where data, analytics and the cloud intersect. The solution lies in combining those factors, adopting a platform for advanced analytics powered by machine learning that has been optimized for the cloud.

1 A FULL-SPECTRUM APPROACH TO CYBER

Cyber threats, which for five years have topped the Director of National Intelligence's list of most dangerous global strategic threats, have never been more severe, varied or prevalent. Malicious actors, whether state-sponsored or independent, have any number of sophisticated tools at their disposal and a wide array of available attack vectors, from phishing emails to communications disruptions. For the Department of Defense and the Intelligence Community, securing information networks and far-flung operations around the world requires a lot more than manpower and firewalls.

Within DoD, the U.S. Cyber Command in the fall announced that its Cyber Mission Forces, comprising 133 separate teams and more than 6,000 service members and civilian employees, was fully operational, but Pentagon leaders have been quick to emphasize the importance of machines in moving forward. DoD classified cyber as an official domain of warfare in 2011, along with land, sea, air and space (and has since mulled adding the electromagnetic spectrum to the list), but the Pentagon has emphasized that cyber defense is more than just a military matter, promoting a "whole of government" approach. One of the three pillars of DoD's Cyber Strategy, along with protecting military networks and missions, covers defending the U.S. infrastructure and national interests.

The Office of the Director of National Intelligence, meanwhile, stresses in its National Counterintelligence Strategy the importance of extending counterintelligence work into the cyber realm, including efforts to "bolster our collection and analytic efforts." In espionage circles, human intelligence – or HUMINT, information gathered by agents in the field – has always carried the most weight, but other sources, such as geospatial, signals, open-source (publically available) and, especially, cyber intelligence have become increasingly important. Those sources produce a lot of data in an array of digital formats, prompting intel leaders to focus on ways to automate analysis, with help from both inside and outside government. The Pentagon's Defense Innovation Unit Experimental (DIUx), for example, has sought to enlist industry in developing new approaches to cybersecurity.

The military gathers massive volumes of data in a variety of formats, and because of the need to quickly turn that data into actionable information, the Cyber Command is "very much interested in artificial intelligence [and] machine learning," Adm. Michael Rogers, commander of the Cyber Command and National Security Agency director, has told the Senate Armed Services Committee. "Because if we're just going to take this largely human capital approach to doing business, that is a losing strategy."

Employing newer technologies such as cloud computing and big data analytics opens up new cyber capabilities for DoD, but it also creates new attack surfaces to worry about. "Cloud computing has given us a new paradigm where we don't have to own everything now," Ron Ross, a fellow at the National Institute of Standards and Technology (NIST) specializing in information security, risk management, and systems security engineering, said in an interview with Cloudera's Cybersecurity on Call podcast and Vision Blog. But while cloud enables the collection and analysis of large stores of data, that data also must be secured. Likewise the Internet of Things, which feeds data stores through any number of connected devices – ranging from heart monitors to power meters – and advanced analytic programs that can sort through all that data and help operators with decision-making. "Those AI programs? Those are subject to attack as well," Ross said.

As new technologies come into the enterprise, Ross said, questions arise about how to secure them. "The answer is, the same controls are applied across all these different technologies," he said. They may be treated in different ways, depending on their functionality and purpose. The IoT, for example, includes a lot of small, low-power devices that work differently than, say, an organization's information network. "But the basic blocking and tackling – the fundamentals of computer security – always apply when there are computers and software and firmware involved," Ross said.

And that's where cloud and advanced analytics can help. Their use might expand the attack surface available to adversaries by increasing the number of endpoints on a network, but they also can be applied to monitoring, pattern recognition and other cybersecurity tactics that keep an environment secure. The big data analytics enabled by those technologies also can help shrink the attack surface by identifying underused assets, redundant processes and other inefficiencies that could be consolidated or eliminated.

One example is a cybersecurity solution such as Cloudera's, which is based on Apache Spot, a community-driven, open source project designed to use machine learning and advanced analytics to help secure systems operating at a cloud scale by finding attacks that might otherwise go undetected. Diligent software updates and regular patching can take care of most known cyber threats, but malicious actors are always coming up with new, unknown – Zero Day – attacks that can get past traditional defenses. Drawing on a wealth of expertise – in addition to Cloudera, Spot's community includes Centrify, Intel, Webroot and others – Apache Spot provides the transparency and tools to help discover patterns that could reveal a previously unseen attack.
Cloudera's solution, for instance, delivers anomaly detection, behavior analytics, and comprehensive enterprise data access on a scalable open platform, and allows customers to customize their solutions.

DoD, like many other organizations, may have trouble developing these kinds of services on their own. A Ponemon Institute research report on the subject found that 65 percent of respondents agreed that cybersecurity big data analytics is critical, but said that a lack of in-house expertise (65 percent), insufficient resources (63 percent) and technology limitations (60 percent) presented significant obstacles. Seventy-two percent of organizations said it was impossible to leverage big data analytics with existing systems.

But organizations recognize the need. The report found that user demand for cybersecurity analytics had increased 71 percent in 12 months, and that cybersecurity analytics made it 2.25 times more likely that organizations would be able to detect threats within minutes.

2 A PLATFORM WHERE DATA LIVES – AND BREATHES

An army may travel on its stomach, as the saying goes, but it lives on its data. Collecting, managing, analyzing and sharing data is at the core of future of DoD operations.
In the battlespace, that covers a lot of ground, from command center views to deployed operations in austere environments. At the center of it for DoD is the Joint Information Environment, or JIE, which the Defense Information Systems Agency, in its five-year Strategic Plan, calls the "the cornerstone of the Department's future." The goal of the JIE is to provide a secure communications infrastructure for joint forces of the U.S. military and allied partners. To get there, the department is pursuing cloud-based solutions that provide access and analytics capabilities across the DoD Information Networks (DoDIN).

Feeding into that big picture are elements ranging from manned and unmanned aircraft that provide millions of hours of full-motion video that overwhelms analysts to other sensors and devices that produce data in a variety of formats, which the department has struggled to handle. DoD last year initiated its Algorithmic Warfare Cross-Function Team, also known as Project Maven, to try to automate analysis of all that video.
The Army Research Laboratory has launched a program, the Internet of Battlefield Things (IOBT), that aims to tap into those data feeds to provide predictive analytics for command and control and battlefield services. And the Defense Advanced Research Projects Agency's SHARE program wants to find a way to process and securely share information across multiple classification levels from a single handheld device.

Meanwhile, DoD's mission also relies heavily on logistics, supply chains, acquisition, health care and human resources, to name a few critical areas. And all of those operations increasingly rely on data management, which DoD leaders acknowledge could stand to be improved.

One example is more efficient acquisition, something DoD has long struggled with. As the Inspector General's report on top management challenges points out, military organizations, preoccupied with short-term costs and performance trade-offs, often pay too much, buy too little and wind up delivering less capability than they need. Acquisition reforms ordered by Congress or the Pentagon have helped in some cases, but many DoD components continue to lag behind. As a solution, the Congressional Research Service has recommended that DoD build a data-driven culture that "integrates data gathering and analysis into the very fabric of the organization."

The supply chain is another weak link. The Government Accountability Office includes Supply Chain Management on its 2017 High Risk Report, and recommends better integration and sharing of data metrics.

Cloud computing can solve a lot of these problems, which is the focus of JIE and DISA's Big Data Platform. DoD's vision for JIE has the ostensibly contradictory goals of expanding its enterprise infrastructure across formerly separate military networks while at the same time reducing the network's cyber attack surface. For that, it uses the Joint Regional Security Stacks, managed by DISA, to consolidate security into centralized regional architectures rather than relying on the perimeter of a globally distributed network. DISA – currently installing 10 of a planned 11 JRSS sites in the continental United States and five planned sites planned outside the country – is working to make greater use of the cloud and automation to take in and process large data sets, and ultimately improve security.

Of course, cloud is not an end in itself. It has its limits, particularly for sensitive systems, such as those handling classified data that must be handled on-premises. Most cloud services stand alone, operating in segregated environments, which can restrict information sharing and complicate security. And the computing environment is made more complex by the addition of more endpoints, such as the myriad devices that can be connected via the Internet of Things. Organizations need a platform that incorporates security, governance and compliance requirements while allowing for a shared data experience at the same time.

Cloudera's Shared Data Experience (SDX), for example, provides a single platform for core functions such as data engineering, data science, and analytic and operational databases. Its unified model can be the most cost-effective and fastest system to deploy while being easiest to secure and govern.

SDX solves the problem of isolated clusters by sharing persistent data and metadata across on-demand applications, ensuring that each cluster doesn't need to be controlled separately. And by centralizing security controls, SDX applies policies and practices consistently across all applications, so they don't have to be reapplied when data is moved or a new application is added into the mix. Cloudera Navigator's shared catalog of data and metadata also makes it easier to meet compliance mandates – whether they fall under NIST's Cybersecurity Framework, the Federal Information Security Management Act, the Health Insurance Portability and Accountability Act or other regulations – and support audits by making short work of finding data and applications, understanding where it came from and tracking how it's been modified.

SDX supports multiple public and private cloud environments, as well as single-tenant, "bare metal" configurations.

3 MAKING THE INTERNET OF THINGS PAY OFF

In almost any future mission facing DoD and its components, a common thread is the Internet of Things. Whether involving monitoring valuable assets, physical security, cybersecurity, health care, maintenance or other tasks, an increasing amount of the data in a data-driven enterprise will come from the IoT.

Much of the IoT consists of connected devices that typically don't have operating systems or much computing power, but can transmit data. And they come in any number of forms – cameras and sensors on unmanned vehicles, smartphones, wearable devices, temperature gauges, water-use meters, light fixtures, anything. The number of IoT devices is vast and only continuing to grow. Gartner estimates that there were 8.4 billion IoT devices worldwide in 2017, and there will be more than 20 billion by 2020. DoD, which has plenty of IoT and other connected devices itself, can expect a similar rate of growth.

The benefits from the IoT can be profound, ranging from more efficient use of resources to a real-time view of assets and better informed decision-making, but it also comes with a set of unique challenges. The most obvious is in analyzing all the data generated by IoT devices and turning it quickly into actionable information. The other challenge is in securing those devices, which can introduce a range of vulnerabilities into the network, as the Government Accountability Office outlines in a July 2017 report.

When properly managed, IoT devices feed into an advanced analytics platform that can, depending on the application, reduce downtime, lower costs and avoid future problems. A good example is with Sikorsky, a division of Lockheed Martin that supplies many of the U.S. military's helicopters. Sikorsky combines IoT sensor data from helicopters with other sources into a scalable data management and analytics platform built on Cloudera Enterprise. Using Cloudera's Enterprise's Data Hub, an Apache Hadoop platform, with a Python data analysis tool, it supports improved aircraft safety and production design, while enabling predictive maintenance. Sikorsky, whose helicopters also are used by first responders and other public safety organizations in addition to DoD, says it set out to improve flight safety and discovered a lot of benefits in the process.

"Safety was the driver, and now we're seeing that result in advantages across the board," said Matt Tarascio, Sikorsky's director of Intelligent Technologies, Analytics, and Sustainment. "Cost reductions, availability increases, readiness increases, having the right parts in the right place, so we're really flowing that out across all of our product line."

Navistar, which builds defense vehicles along with commercial trucks, buses and engines, found similar results when it used Cloudera Enterprise to build its remote diagnostics platform, which it calls OnCommand Connection. For most vehicles, scheduled maintenance is based on one of two things: distance traveled (like the light on the dashboard of your car, which is tied to the odometer) and time elapsed since previous maintenance work. Even warnings such as a "check engine" light only provide a clue about potential problems, and in any case, they're reacting to something that's already happened.

Navistar's Cloudera platform taps into a wide range of data points that deliver real-time information and, combined with diagnostics and analytics, can predict when maintenance is necessary before something goes wrong. "We can evaluate billions of rows of data from connected vehicles in hours, not weeks, to enable predictive maintenance," Navistar CIO Terry Kline said.

The platform, which can be monitored from smartphones or tablets, combines information from more than 375,000 connected vehicles (including engine performance, truck speed, acceleration, coolant temperature and brake wear) with other data such as meteorological, geolocation, usage, traffic and parts inventory information. Using machine learning and advanced analytics, it predict failures such as looming engine problems, even locate the nearest dealer with the necessary parts in stock and identify available technicians and service bays. The end result is something DoD components can appreciate: reduced downtime and lower maintenance costs. Not only is the IoT growing, but the data from each connected device is soon to grow as well. The Defense Microelectronics Activity, for instance, is in the early stages of a 10-year, $8 billion program to upgrade DoD's embedded microelectronics with advanced technologies. And DARPA recently launched a project with industry and academia to address the next wave of microelectronics technologies, including adding artificial intelligence capabilities and much faster data transfer rates to IoT devices, which will only increase the amount of incoming data – and make more clear the need for a platform to handle data analytics.

4 A PRESCRIPTION FOR BETTER HEALTH CARE

The Military Health System (MHS) has a simple, if daunting, mandate: Provide top-notch, affordable care to about 10 million beneficiaries around the world – active-duty personnel, retirees and their families –while keeping pace with new technologies, new treatments, fluctuating policies and standards for access, and evolving client expectations. Beyond treatment of injuries and illnesses, DoD also is looking to use technology for preventive care, monitoring personnel to reduce the risk of physical or psychological injuries.

Two examples of the challenges in the broad swath of services can be seen at seemingly opposite ends of the data management and analytics spectrum: electronic health care records dating back years in various formats in various systems, and real-time monitoring of military personnel and patients via wearable devices and other machines connected to the Internet of Things.

Creating a unified health record has proved to be a Sisyphean task for DoD and VA since 1998, when they set out to create interoperable records to follow service members in the transition from active duty to veteran status. The current goal of the DoD and VA Information Exchange is a kind of cradle-to-grave health record, starting with their entry into the Military Health System and continuing through subsequent updates and eventually on to their retirement and care under VA. During one period, the two organizations spent four years and $1 billion on a joint records program called iEHR, only to abandon it in 2013 in favor of separate systems. After a few other fits and starts, the two agencies are coming back together. VA announced in June 2017 that it would use DoD's MHS Genesis system for electronic health records, which is being built under a 10-year contract awarded in 2015 and projected to cost $10 billion.

The centerpiece of MHS Genesis is Cerner's Millennium platform, which provides the core capability as a hosted software-as-a-service. DoD has said the company's data center allows direct access to proprietary data that it couldn't get from a government-hosted environment. And Cerner has experience in dealing with the massive volume of data created by a health system that serves 10 million beneficiaries at hospitals and clinics around the world.

Cerner uses a modern platform powered by Cloudera. The companies began working together in 2010, when Cerner

adopted Cloudera's open source platform distribution which allows random access to data. When Cerner, whose services are used at more than 14,000 locations around the world, set out to create a comprehensive population health platform and needed a way to employee significant computational power while maintaining flexibility in handling its records, it chose Cloudera again.

With Cloudera, the platform takes in data with various formats from multiple sources – electronic medical records (EMRs), Health Level 7 International (HL7) feeds, Health Information Exchange (HIE) information, claims data, and extracts from proprietary or client-owned systems. The platform can absorb data streams in real time, then pass them on to the right database or distributed file system. With Cloudera, Cerner manages more than 2 petabytes of data in support of both patient and financial matters for several hundred clients.

Cloudera "provides a holistic view of our whole environment, and allows us to manage multiple clusters from a central point," said David Edwards, Vice President and Fellow at Cerner. Collecting and analyzing data from almost unlimited sources allows clients to create a far more complete picture of patients, their conditions and treatments. "We're able to achieve much better outcomes, both patient-related and financial, than we ever could by just looking at pieces of the puzzle individually," said Ryan Brush, Senior Director and Distinguished Engineer for Cerner.

Cerner also is using Cloudera tools to perform deep data science and build predictive models that can determine the likelihood that a patient could be readmitted with the same condition or something similar, or whether they could develop other complications, such as a bloodstream infection. "Our clients are reporting that the new system has actually saved hundreds of lives by being able to predict if a patient is septic more effectively than they could before," Brush said.
Edwards said Cerner's ultimate goal is to have technology that is effective and seamless enough to operate almost unseen, allowing medical professionals to only apply the fruits of that technology to providing better medical care.

On a different front is the IoT. Machine learning systems have demonstrated a propensity for improving care in a number of areas that are of interest to DoD, such as behavioral health (including PTSD), suicide prevention, diagnoses, disease research and treatment. DoD is investing in a variety of biosensors and monitoring devices, including smart fabrics and wearable devices to track fitness, general health and sleep, and stress and fatigue levels. The Air Force Research Laboratory is developing AI-based "synthetic partners" that would one day assist pilots and other airmen in their duties while simultaneously monitoring vital signs, stress levels and other factors.

Combined with machine learning, the IoT is expected to see enormous growth in the medical field over the next decade. But it already is being used extensively now.

One example is Cerner's efforts to ward off sepsis, which affects as many as 18 million people around the world each year. The company's IoT electronic warning system, the St. John's Sepsis Agent, remotely monitors patient's symptoms and sends data to a central hub where advanced algorithms can predict an increased level of risk.

Another is DocBox, a healthcare IT company that uses IoT medical devices to automatically collect the full set of a patient's vital signs and waveforms, which in an intensive care unit can come from as many as 300 sources. As a result, nurses spend less time transcribing data and more time delivering care, while the analytics system processes information

and presents it in an easily digestible format, allowing medical professionals to note any subtle changes in a patient's condition, or personalize a treatment plan.

A national children's hospital, meanwhile, also is demonstrating how the IoT and analytics can be used to improve other elements of patient care, specifically with regard to infants. The hospital collected respiration, heart rate, blood pressure and other data to study the effects that noise and light levels had on babies, analyzing massive data sets to find ways of improving care. The Cloudera platform accumulated 2 terabytes of data in the first few months, and has been adding about 50 gigabytes per week since.

5 THE TOOLS OF MODERNIZATION

Streamlining operations is a permanent goal of just about any organization, and DoD's leadership has turned up the heat. Secretary of Defense Jim Mattis has established cross-functional teams to look for efficiencies and has pushed for enterprise-wide consolidation of business activities including human resources, financial management, acquisitions, logistics, health care and cyber operations.

Meeting those mandates, while also achieving mission goals, requires modernization across the board, whether in communications, cybersecurity, financial management, health care or other areas. And for many of DoD's most urgent needs, modernization will involve two key elements – a cloud infrastructure and greater automation.

Data center consolidation, cross-domain information sharing, improved cybersecurity and big data analytics – not to mention budget requirements – all benefit from a cloud infrastructure, which in turn feeds the analytics that DoD has deemed essential to future operations. Some of DoD's efforts in those areas aren't quite succeeding on their own.

The department's struggle with closing and consolidating its data centers is one example. The Data Center Optimization Initiative (DCOI) and the Federal Information Technology Acquisition Reform Act (FITARA) has led to the closure of thousands of government data centers, but the initiative has stalled. The Army, for instance, had planned to shut down 60 percent of its roughly 1,200 data centers by this fiscal year, but will fall short of that goal, and DoD overall received a grade of F for data center optimization on the latest FITARA scorecard. Dave Powner, director of IT issues at the Government Accountability Office, has suggested that agencies might have to turn over the job to someone else. "If an agency really can't optimize by 2020, should they be in the business of managing a data center?" Powner said at a data center conference last year. "They need to ask themselves that. If agencies can't operate these things, they need to think long and hard about getting out of the business."

The trend line for data center workloads is continuing to point to the cloud. Cloud-based data center traffic worldwide was already at 88 percent in 2016, and is expected to hit 95 percent by 2021, according to Cisco's latest Global Cloud Index. The volume of that data will more than triple in that same timeframe, going from 6 zettabytes in 2016 to 19.5 zettabytes in 2021, with cloud applications and the IoT fueling much of that growth. The top growth areas among cloud applications for organizations? Enterprise resource planning, collaboration, and analytics, according to the report.

And because next-generation cloud and analytics technologies both require and generate a lot of data, an essential focus of modernization is with an organization's enterprise data warehouse (EDW). An EDW is the central repository for any and all data an organization wants to analyze, which in DoD's case involves a monumental amount of data, in varying formats, that needs to be tracked and quickly analyzed to support better decision-making.

An effective approach is to incorporate an analytics database, where the EDW can offload extract-transform-load (ETL), business intelligence (BI) and analytic workloads in order to make the most use of its data while maintaining its existing footprint. The Cloudera analytic database applies high-performance SQL analytics to data in a cloud-native environment, allowing analysis of data wherever it may be, without having to move or copy it.

Cloudera's modern analytic database, powered by Apache Impala (incubating), also offers an open architecture, elastic scalability and seamless data portability features that are unique to cloud-based analytic DBs. Impala is a massively parallel processing (MPP) engine native to Hadoop that supports multiple workloads simultaneously, giving analysts access across to data across the enterprise and quickly delivering insights. Impala's hybrid portability is built to work with data stored on open, shared platforms, enabling access to data from multiple sources in open formats. That flexibility lets users avoid the type of rigid data modeling and loading that's found in most monolithic analytic database architectures.

By taking the load off of an EDW and integrating with it to allow for curated reporting, the analytic database and its elastic scalability – provisioning resources to match demand at any given moment – lets organizations integrate data faster, and provides a lot of flexibility, allowing more self-service analytics and reporting for users. With this strategy, the database can accommodate multi-tenant environments across all shared data. The scalability and optimization also deliver a high degree of cost-effectiveness, which is a crucial factor for any organization.

As with many of its other initiatives, DoD's enterprise modernization efforts will ride on the cloud. A open platform that provides a flexible, interoperable approach with nearly unlimited capacity can help ensure that a cloud-based environment will not only meet current needs but be able to adapt to future demands.

CONCLUSION

The Army defines actionable intelligence as "high level of shared situational understanding, delivered with the speed, accuracy, and timeliness necessary [for commanders and soldiers] to operate at their highest potential and conduct successful operations." One of the key elements in creating actionable intelligence is "an intelligence framework linking analytic centers, databases, and sensors."

The Army's definition focuses specifically on the battlefield, but the same concept can be seamlessly applied to any field that uses data and analytics to glean useful information. Medical data that indicates the likelihood of a reoccurrence of symptoms is actionable. Pattern recognition that uncovers anomalous network behavior consistent with cyberattacks are actionable. So are readouts from a predictive analytics system that tells you a key helicopter component will fail before it actually fails.

The ability to turn that data into something actionable relies on a framework that can combine data from records, videos, the Internet of Things and other sources – delivered via the cloud – analyze it quickly and autonomously, and present its conclusions clearly to users, be they soldiers, doctors, managers or scientists.

DoD, like organizations in virtually any sector, has hit an inflection point with cloud computing. Regardless of how many systems operate in the cloud now, it is both the present and future. The demands for secure information sharing and advanced data analytics – on ever-expanding data sets fed by cloud-based systems and the IoT – requires an agile, comprehensive and cost-effective approach that can scale to meet future needs and incorporate new technologies. Actionable intelligence depends on it.