BIG DATA PLATFORM WITH SECURITY VAULT ENABLES REGULATORY COMPLIANCE
Big data has moved from being an over-hyped term to yielding significant gains for organizations through steadily increasing adoption of meaningful, ROI-based use cases. CIOs are focused on gaining clarity on what's driving and shaping big data initiatives.

At this time, there is no single big data app that you can take off the shelf, install and operate meaningfully and profitably. As big data and analytics become progressively embedded in the fabric of business operations, security and data protection concerns come to the fore.

Regulation, technology standards and approaches are only now emerging. The space is changing rapidly, with new components arriving every day — which makes it very difficult to standardize the architecture, integrate technologies, define solutions, ensure enterprise-grade security and hire a team with the knowledge to meet your business needs.

When people use the term “big data,” they are referring to a combination of infrastructures, applications, data sources, analytical models and data repositories, sometimes on an enormous scale — all of which makes securing big data a unique challenge. This paper outlines some of the key aspects essential to ensuring a secure and robust big data platform for the enterprise. It examines areas where CSC’s solution goes beyond the standard in maintaining enterprise-grade data security and integrity, making CSC’s Big Data Platform as a Service (BDPaaS) the best option for the enterprise.

DATA PROTECTION
Data is the crown jewel of any business, and needs to be protected from theft and loss.

At a time when data is being collected on people, places and things at an exponential rate, we should be worried about putting that data at risk. Cyberthreats reported in the news today lead us to a heightened concern over the security of the organization and its mission-critical business data. An organization that intends to collect and store data is a ripe target for industrial espionage, hackers and criminal elements, to name just a few.

Data varies in value based on its sensitivity in the business operation. Patient record data, for example, needs to be treated differently from staff vacation data. Therefore, the data protection mechanisms must also differ, and must cover the logical and physical security of the data center, the infrastructure, support staff and company processes. The risk profile starts to grow exponentially when emerging and open source technologies are added.

DATA BREACHES
Three types of data breaches affect organizations today: cyberattacks, insider threats and unintentional data loss.

• Cyberattacks are targeted, planned and usually meant to steal payment card data, customer information, trade secrets and other IP, damage brands, or commit further illegal acts with the stolen data.
• Insider threats have always been there, but with the proliferation of data assets that affect companies’ bottom lines, they are becoming more prevalent.
• Many issues can lead to unintentional data loss or compromise: disasters, configuration errors, corruption, or poor data-handling practices stemming from unintentional actions. All of these can hurt your company’s reputation, and in some cases have led to bankruptcy filing, fines or company closures.

THREE ASPECTS OF DATA SECURITY AND INTEGRITY
The key to using data effectively and being able to leverage it for advanced analytics is to have the right people, processes and technologies in place; the same as any IT solution. It requires comprehensive attention not only to securing the logical but also the physical environment. Our solution is only as secure as the weakest of all these elements.
Technology. Whether it’s the disposition of stored data, data moving between nodes, or data in active use inside a process, we must use a variety of technologies to secure that information, ensure its integrity and prevent its loss. Data encryption, audit log management, antivirus tools, infrastructure and network security-monitoring solutions to detect sensitive assets moving across the network and flagging unauthorized behavior are just some of the elements that make for a secure solution.

People. Technology may constitute the traditional focus of a secure solution, but it is the human aspect that often is the most problematic. Security-conscious employees who understand the process and act with integrity are paramount for adequately defending against and preventing security threats. In addition, these people must work in the right locations with the right workplace security. Infrastructure and monitoring tools will also flag unusual employee activity (e.g., a system administrator making unauthorized edits to sensitive asset storage, or an employee moving 100,000 customer records, when they typically move only 2,000 at a time) to protect against insider security threats. Tight restriction of an organization’s most sensitive assets can also be achieved through identity and access management (IAM) technologies to lock down authentication to the network, based on policy. The IAM technologies would then give authorization to users based on their “entitlements,” also as defined by policy.

Process. Problems rarely follow an exact script, but having an incident response plan to ensure processes and governance are in place to manage a range of scenarios is crucial. From internal issues related to load, infrastructure or software failures to outside issues ranging from cyberattacks to natural disasters, every secure solution requires a range of triggers that alert the staff to follow certain steps. Such procedures ensure that data will be secure, intact and available to meet a client’s service level agreements. Everything from the thoroughness of our compliance checklists to a monitoring and audit logging solution plays a key role in policy process definition and enforcement. An incident response plan dramatically reduces the time to remediate and resolve whatever caused the outage, whether it was a cyberattack or destruction from a hurricane. Those who act fast have the best chance of recovering their business. We have the capability today to put plans in place to know who to notify in case of an outage to meet customer service, business requirements and industry regulations.
As shown in the figure, CSC BDPaaS is a fully integrated and managed big data environment, delivered as a service, that enables clients to rapidly acquire and apply analytics insights through batch analytics, fine-grained and interactive analytics, and real-time streaming analytics. CSC BDPaaS makes it faster, simpler and less costly to develop, deploy and manage big data applications securely.

Security is the cornerstone of CSC BDPaaS’s architecture. The solution combines an open source core supplemented with proprietary solutions to support the specific security and integrity requirements of various industries and regulations.

PERIMETER SECURITY
A strong security implementation is vitally important in big data applications, and it starts at the perimeter. Traditional application environments offer fine-grained security and access control. However, big data environments are different, since the applications are tightly interwoven with an operating system and leverage many of its core components. This makes it difficult to isolate the software from the core operations of the infrastructure. Outside, the system is a black box. Once access is gained, a user has the “keys to the kingdom.”

Perimeter security makes it very difficult for an unauthorized outsider to get those keys and gain access. CSC BDPaaS ensures perimeter security through a combination of secure VPN connections, isolated subnets, secure port management, fine-grained port monitoring, IP address whitelisting and blacklisting, masquerading all Hadoop services behind a Knox Gateway, and multiple-level ACL implementations, which include running a big data cluster on a virtual private network, closing ports that allow Web API access, secure authentication and authorization to gain access.

USER MANAGEMENT
User management is achieved by integrating a big data solution with an existing LDAP repository or Active Directory. Frequently, Kerberos is used as a network authentication protocol. This is an area where add-on technologies provide features and control. CSC uses FreeIPA with two-factor authentication as an additional component to ensure an open source framework for centralized user management. Access to Hadoop-specific services like Impala, Hive, HBase and HDFS is protected with role-based access control (RBAC). This ensures a single interface that is abstracted, but one that has comprehensive control over the system. CSC BDPaaS architecture is very flexible and can quickly integrate with additional tools such as Apache Knox and Apache Sentry to meet higher levels of user management.
DISK AND NETWORK ENCRYPTION

Data encryption in CSC BDPaaS addresses an area in big data that has been ignored for too long. In the latest Hadoop release, more attention is being paid to data encryption between nodes and for data in use in processes. Many new elements of data encryption, such as encrypted node-to-node communication, are now appearing in the core Hadoop solution through acquisitions (Cloudera buying Gazzang, Hortonworks buying XA Secure). These elements strengthen the encryption of data in the Hadoop ecosystem.

CSC further leverages Kerberos and tools from operating systems (for example, Red Hat LUKS) for disk and wire encryption. Certified partners are brought in when additional security and compliance standards are needed. CSC maintains partnerships with selected vendors in data encryption and chooses the most appropriate add-ons for clients as part of our solution. For example, Zettaset encryption modules can be used for encryption of data at rest and data in motion — an important requirement for HIPAA compliance. Amazon Simple Storage Service (S3) encryption and Amazon Elastic Block Store (EBS) volume encryption are additional options for disk and network encryption for Amazon Web Services (AWS).

ACTIVITY MONITORING

CSC BDPaaS facilitates log tracking of all events in the system through Splunk. Log tracking includes documenting changes, logins/logouts, jobs runs, queries and indexing activities to ease searching by compliance or audit teams. A combination of Puppet, Cloudera Manager and Ambari is used to manage and track configuration information across the platform. Add-on modules are configured to make reporting easier when required for regulatory compliance. For example, HP ArcSight Security Information and Event Management (SIEM) can be used for advanced security event management and audit logging. The tools used for activity logging also extend to monitor system health and manage the system’s key performance indicators (KPIs): system uptime, utilization and performance. These tools and the related data are used to drive dashboards and generate reports.

HARDENED OPERATING ENVIRONMENT

Standard operating environments are not just about deploying an operating system (OS), they are about deploying a complete platform stack, with system management tools and data collection tools with baseline mapping. For a secure, enterprise-grade big data platform, the operating system needs to be both virtualization ready and antivirus ready. Vulnerability scanning and penetration testing for every major software release, security patches applied to both the core product as well as to products used for customer solutions, source code scanning for vulnerabilities, and operating system patch application are all standard operating procedures for BDPaaS. CSC Endpoint Security (TrendMicro) for hardware, ClamAV for virtual machines, and Tripwire for monitoring file integrity are some of the other features that ensure BDPaaS security.

INFRASTRUCTURE/CLOUD SECURITY

The infrastructure of a big data stack is just as important as operating systems and applications in matters of security and integrity. Big data infrastructure may be a dedicated set of hardware assembled by CSC and located in a CSC data center or in customer data centers. However, the largest share of big data engagements today is cloud-based. CSC’s BDPaaS can be hosted on any cloud, but most often is operated on either CSC’s cloud or AWS. CSC’s Cloud Managed Security Services enable the most efficient and “future proof” transformation of an enterprise security state, with the ability to continually adapt as business needs and associated risks change. CSC Cybersecurity has been recognized as a leader in the “IDC MarketScape: Worldwide Managed Security Services 2014 Vendor Assessment” for our capability in global managed security services and innovative delivery model with extremely high customer satisfaction.
On the other end, Amazon has been trusted in multiple high-security environments, including the National Security Administration (NSA). In addition, AWS has strong familiarity with handling ISO requirements.

**ADDRESSING REGULATORY AND COMPLIANCE REQUIREMENTS**

Regulatory standards and compliance requirements further influence the three aspects of people, process and technology in securing a big data platform. There is a lot of similarity between regulatory requirements across industries, as many are designed from existing government standards and modified to meet the needs of specific data sets used in an industry.

However, standards in one industry may be more stringent in the means used to secure data in storage, while regulations in another industry might be more concerned about data movement between nodes. These differences extend to people and processes as well. For example, it’s possible that a delivery center certified as HIPAA-compliant may not be Payment Card Industry (PCI)-compliant at the same time. Documentation differs between standards, and the methods for handling data can vary widely between patient-identifiable data under HIPAA standards and credit card data managed under PCI standards. BDPaaS is architected and tested to support a variety of regulatory and compliance requirements.

**DATA GOVERNANCE**

Security governance for CSC BDPaaS is ensured through implementation of an information security management system (ISMS) at the start of the project to incorporate key stakeholders and decision makers. The ISMS is developed against ISO 27001 standards. Controls are developed using the Control Objectives for Information and Related Technology (COBIT) framework, as COBIT is commonly used by traditional auditors for internal and external audits. All the objectives, guidelines and frameworks are condensed into a single control and compliance framework that encompasses all of the relevant standards and IT audit objectives. This creates a unified compliance program and a comprehensive approach to information security.

**PROTECTING DATA INTEGRITY**

Data security is about more than deliberate attacks and theft prevention. The actual availability of data is critical. A big data managed service engagement must include architectural approaches and technologies that support high availability and disaster recovery. CSC’s BDPaaS embeds features to ensure high availability and integrity of customer data. Service level agreements (SLAs) for data availability are driven by specific business requirements, and vary across service availability/response time, recovery time objectives and recovery point objectives. Multiple high-availability and disaster recovery systems can also be put in place to meet different needs, depending on the value of data used in a specific environment.

**SECURITY AND COMPLIANCE AT THE APPLICATION AND ANALYTICS LAYER**

The huge amount of flexibility that Hadoop and other data toolsets provide developers and analysts also brings considerable security risks at the application layer. One of the ways to help build more fine-grained controls around application and analytics data is to manage data at the event and interaction level. Hadoop projects such as Apache Ranger, Apache Knox and Cloudera Sentry all provide frameworks for protecting API, query and data interactions with the system.

CSC’s business consulting and data science services assist by recommending and using appropriate tools for data management at the application and analytics layer, where derivative data is being produced, data is being moved around, and data is being exposed through interfaces such as machine-to-machine data flows, business intelligence and data visualization tools.
Application security testing is important to determine if the application has been compromised or is vulnerable to unauthorized access. Cybersecurity Application Security Testing provides testing and apps modernization processes to ensure the application remains secure. CSC Cybersecurity’s Advanced Threat Detection Managed Security Service continuously monitors networks for anomalous behavior, indicating a possible breach.

PHYSICAL SECURITY
The human aspect of security — in the form of adequate controls in place, continual training, best practices implementation and other compliance-readiness activities — is equally essential. Some additional steps at CSC include:

- Employee background checks
- Employee undertakings, which cover information security clauses and confidentiality agreements specific to clients
- Training to comply with copyright, code of ethics, etc.
- Continuous information security awareness training
- Locked-down physical facilities with active monitoring

CONCLUSION
To sum it up, CSC BDPaaS simplifies the big data landscape by wrapping logical and physical perimeter security, user management, encryption, alert management, activity monitoring, audit logging, and all the other necessary security technologies around an open source core. BDPaaS is supported by people and security processes that offer clients the biggest benefits of big data open source technologies combined with the security features of proprietary solutions.

ABOUT CSC BDPAAAS
CSC BDPaaS is a fully integrated and managed big data environment, delivered as a service, to rapidly acquire and apply analytics insights through batch analytics, interactive analytics and real-time streaming analytics. CSC BDPaaS makes it faster, simpler and far less costly to build and manage big data applications. With CSC BDPaaS, enterprises can better understand customer behavior, product innovation, risks and operational efficiencies through predictive and prescriptive analytics of internal and external data.

Learn more at csc.com/bigdata. Contact us at csc.com/contact_us to request a demo or for more information.
Regional CSC Headquarters

The Americas
3170 Fairview Park Drive
Falls Church, Virginia 22042
United States
+1.703.876.1000

Asia, Middle East, Africa
Level 9, UE BizHub East
6 Changi Business Park Avenue 1
Singapore 468017
Republic of Singapore
+65.6809.9000

Australia
26 Talavera Road
Macquarie Park, NSW 2113
Australia
+61(2)9034.3000

Central and Eastern Europe
Abraham-Lincoln-Park 1
65189 Wiesbaden
Germany
+49.611.1420

Nordic and Baltic Region
Retortvej 8
DK-2500 Valby
Denmark
+45.36.14.4000

South and West Europe
Immeuble Balzac
10 place des Vosges
92072 Paris la Défense Cedex
France
+33.1.55.707070

UK, Ireland and Netherlands
Floor 4
One Pancras Square
London
NIC 4AG
United Kingdom
+44.020.3696.3000

About CSC
CSC is a global leader in next-generation IT services and solutions. The company’s mission is to enable superior returns on our clients’ technology investments through best-in-class industry solutions, domain expertise and global scale. For more information, visit us at www.csc.com.