CYBERSECURITY - FROM EDGE TO AI

Cloudera Cybersecurity Platform

Incorporate existing SOC investments and assimilate new technologies, sources, data and ML models over time – and in real-time. In your data center or in the cloud. Or both.

Protect The Business

$450 Billion: Annual Impact of Cyber Incidents

67%: Enterprises Exposed to Cyber threats

8 months: Average Time an Advanced Security Breach Goes Unnoticed

82%: Percentage of Breaches that Happened in Minutes

41%: Breaches Not Investigated

26%: Investigated, but Not Remediated

Increase SOC Team Agility

- Ingest telemetry data from any security endpoint service, log, intrusion detection system and network/threat intel feed agent.
- Monitor & detect cyber anomalies in realtime.
- Store & analyze long term logs for detecting previously unseen APTs earlier in the kill chain

Bring Machine Learning Closer to Your Front Lines

- Leverage both the latest and most established Data Science and Machine Learning models.
- No need to sample. Run analysis & modeling directly on the real data with secure SIEM/SOC clusters.
- Discover reproducible, collaborative research faster. Bring insights to your team sooner for increased agility.
- Give your data science team the freedom to work how they want, when they want. Securely.
- Stay compliant with out-of-the-box support for complete security.
- Flexible deployment. Run on-premises or in the cloud.

Why Cloudera? Why Now?

Cybersecurity has become an urgent topic of conversation for organizations across every industry, and a priority investment among most IT departments. And for good reason: **theft, fraud, lost intellectual property are a persistent threat. Reputational damage alone can even kill a business outright.** As a result, it should not be a surprise that organizations are looking for new ways to detect and investigate cyber threats.

Attackers have become more sophisticated and the attack surfaces that can be exploited have expanded. **As the number of attacks have increased, organizations find themselves exposed to an onslaught of novel and traditional attacks.** Combined with the threat of inside rogue users and limited availability of skilled resources for detecting and responding to these threats, it is clear organizations face an enormous challenge. The disparate and expanding choice of tools available to the Security Operations Center (SOC) are not built for the hyperconnected world they now operate within.

The threat landscape is changing rapidly. **The number of touch points is exploding and so is the number of entry points for malicious activity.** Hackers are getting more sophisticated. Activists are getting more aggressive. Agencies are getting more assertive. With traditional Security Information Events Management (SIEM) applications, organizations face data and analytic constraints that cause threats to go unnoticed and data breaches to happen.

Traditional SIEM ecosystems **cannot monitor every corner of the enterprise because of technology, human resource and economic constraints;** they are hard-pressed to discover known threats until it is too late, and they only hold a subset of data that makes it difficult to use historic data for investigation and remediation. **Most organizations continue to struggle** to leverage advanced analytics on traditional systems to discover advanced threats, this is forcing organizations to rethink their cybersecurity strategy.

That is where the open source software ecosystem and Cloudera come into play. **Open source provides better weapons to the defense of your enterprise, through transparency and community.** The strong governance process of open source projects, with many eyes reviewing the code, removes risk, guaranteeing releases with the kind of screening, testing, and security analysis you could expect from a commercial software vendor. And removes the risk of vendor lock-in while ensuring agility and flexibility as needs evolve.

Sharing code, ideas, and intelligence data is common on the black hat side through dark web communities and market places. **White hat open source communities apply the same kind of community leverage by sharing code and intel, but do so in a community of trust with a watchful eye.**

Cloudera Cybersecurity Platform

- Create a Multi-dimensional View of Risk with a canonical, open source cybersecurity data model
- Recognize, aggregate and standardize the most common cyber source inputs for more complete threat surface area coverage.
- Bring analysis directly to the data.
- Stay compliant with out-of-the-box governance for complete security.
- Flexible deployment. Run on-premises or in the cloud.

Cloudera Data Science Workbench

- Leverage both the latest and most established Data Science and Machine Learning models.
- Give your data science team the freedom to work how they want, when they want.
  Secure by default.
- No need to sample. Directly access data in secure SIEM/SOC clusters.
- Discover reproducible, collaborative research. Share insights with your whole team.

Cloudera Data Flow

- Ingest telemetry data from any security endpoint service, log, intrusion detection system and network/threat intel feed agent.
- Monitor & detect cyber anomalies in realtime.
- Bring in and establish monitoring for any Edge or IoT device or source.

Cloudera Data Platform

- Store & analyze logs for better time series long term detection.
- Identify and prevent previously unseen APTs earlier in the kill chain.
- No need to sample. Directly access data in secure SIEM/SOC clusters.
- Give your data science team the freedom to work how they want, when they want.

Cybersecurity is a **Big Data** Problem. We Have a Big Data **Solution.**

Cloudera Cybersecurity Platform (CCP) modernizes an organization's cybersecurity architecture by leveraging established, high performance, scalability and reliability (PSR) open source big data technologies with Apache Metron operating at its core. Apache Metron is **a highly scalable advanced security analytics framework built with the open source community evolving from the Cisco OpenSOC Project.** Metron provides a canonical data model that offers organizations the ability to detect cyber anomalies in real-time, focused heavily on streaming data and fast data processing at scale to enable organizations to rapidly respond to identified anomalies in telemetry data from most known security endpoint services, machine generated logs, intrusion detection systems and network & threat intel feed source agents, as well as more traditional enterprise transaction systems, such as ERP, CRM, HCM applications. **The result is the ability to detect advanced threats 2.25 times faster(1) and accelerate threat mitigation leveraging big data and advanced analytics** (machine learning, predictive analytics, etc.).

Unlike traditional solutions that provide signature and correlation analysis across subsets of security data, the open source-based **CCP can ingest, store, process, and analyze any volume of data with any analytic type. Having access to all the raw data in one place can help uncover new insights and patterns.** This allows for behavior-driven analytics that can detect the smallest changes in user or system behavior—traditionally the most reliable indicators of compromise.

Integrating existing cyber defenses, CCP allows organizations to quickly deploy and improve their security posture with no disruption.

(1) Source: Ponemon big-data-cybersecurity-analytics-research-report 2016

Cover The Edge. Streaming Realtime, Into The SOC

Edge data streaming, collection and management introduces new threat surface challenges. One of the key challenges that SOC teams are facing today includes how to properly identify, monitor and secure IoT devices and related edge data sources that may connect to the enterprise. These can now be addressed more easily with Cloudera DataFlow (CDF). CDF is a comprehensive edge-to-enterprise streaming data platform that can be incorporated into the CCP ecosystem. It addresses the key data management challenges with streaming and IoT data for all types of enterprises. Cloudera Edge Management (CEM) is a key part of the CDF platform and it addresses IoT and edge data management challenges around data collection and processing edge data from a wide range of edge devices and streaming sources.

CDF manages, controls and monitors edge agents to collect data from edge devices and push intelligence directly to the SOC as well as back to the edge. This allows you to develop, deploy, run and monitor potential security threats across edge flow apps on thousands of edge devices.

CDF is an edge management agent that implements the core features of Apache NiFi, and focuses on data collection & processing at the edge. This enables you to bring more of the edge into view in CCP for the SOC and to apply more comprehensive cyber monitoring across more of your enterprise ecosystem. Coupled with Data Science/ML, a truly multi-dimensional threat surface area can be monitored and defended.

Bring Data Science & Machine Learning Closer to the Front Lines

With the move towards machine learning and artificial intelligence on both sides of the battle, the open source debate takes on a new angle. Powerful platforms from open source software are available to both sides, but **the quality and efficacy of models is usually dictated by data availability for training.** While the attackers lack the scruples to obey compliance and data privacy rules, and can therefore exploit more of the information, they are reliant on information leaked or stolen, while the **defense can usually benefit from a more complete picture if they can use the full range of their data effectively to build defensive models.**

The Cloudera Data Science Workbench (CDSW) is an enterprise data science platform that **accelerates data science and machine learning projects by providing a robust yet familiar environment for model building with self-service access to data.** It provides a complete solution for data exploration, analysis, visualization, modeling and model deployment. **CDSW makes secure, collaborative data science at scale a reality** for the enterprise, accelerating the delivery of new data products and unifying IT, data science, developers, analysts and business stakeholders across the data science workflow.

Data scientists can **deploy trusted machine learning models faster by bringing their tools directly to the data**, whether it's stored in CDH, HDP, CDP. Either on-premise or the cloud. Using a choice of languages including Python, R or Scala directly from the web browser, CDSW delivers a rich, self-service experience for data scientists.

Download the latest libraries and frameworks in customizable project environments. Beyond the Python and R ecosystems, as open data science expands to deep learning frameworks like TensorFlow, PyTorch, Ca e 2, MXNet, DL4J, BigDL, scikit-learn, and more, **CDSW delivers a safe, secure environment to combine the latest open source innovations with the unified platform** Cloudera customers trust.

Collaborative, shareable project environments ensure **diverse data science teams can work together toward standard, reproducible research and production ready models**, and easily deploy and manage them across stakeholders and end-users.

IT groups often struggle to onboard data scientists to big data systems because of their diverse needs, especially where open source tools are involved. The result is duplication and analytic silos with limited security and governance. Meanwhile, data scientists look to scale their work to larger data sets and more powerful compute platforms. **CDSW helps you get better use of your core data management investments by removing analytic silos and drive more value** from your enterprise data platform, whether on-premises or in the public cloud.

Use Long-time Series to Detect & Mitigate the Bigger Risks

The Cloudera Data Platform is built on the latest open source projects and available in a variety of form factors. The result is a well established and proven highly performant, scalable and reliable platform designed to process and store a very large variety, volume and velocity of data that can solve for the most complex and demanding enterprise data management requirements. This delivers a new class of cybersecurity solution – one designed as a modular framework that can both incorporate existing investments and assimilate new technologies, sources, data and ML models over time – and in real-time - to detect previously unseen threats early in the kill chain—helping organizations avoid financial and reputational damage.

Example High Level CCP Architecture

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Cloudera, Inc. 395 Page Mill Road Palo Alto, CA 94306 USA cloudera.com

© 2019 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice. Cloudera_Solution-Brief_Cybersecurity_101