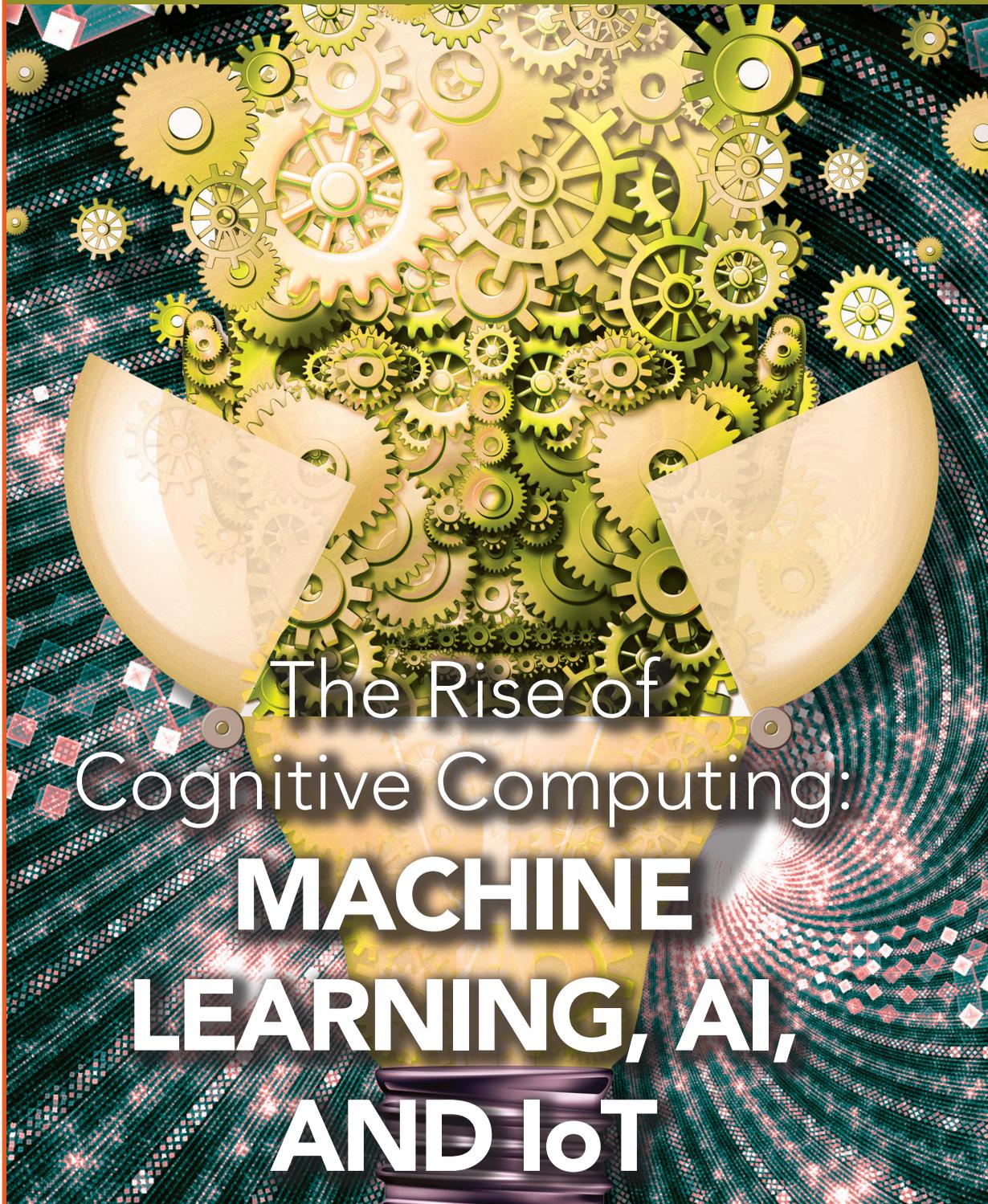


Cloudera

ACCELERATE  
ENTERPRISE MACHINE  
LEARNING WITH  
CLOUDERA

# THOUGHT LEADERSHIP SERIES



## The Rise of Cognitive Computing: **MACHINE LEARNING, AI, AND IoT**

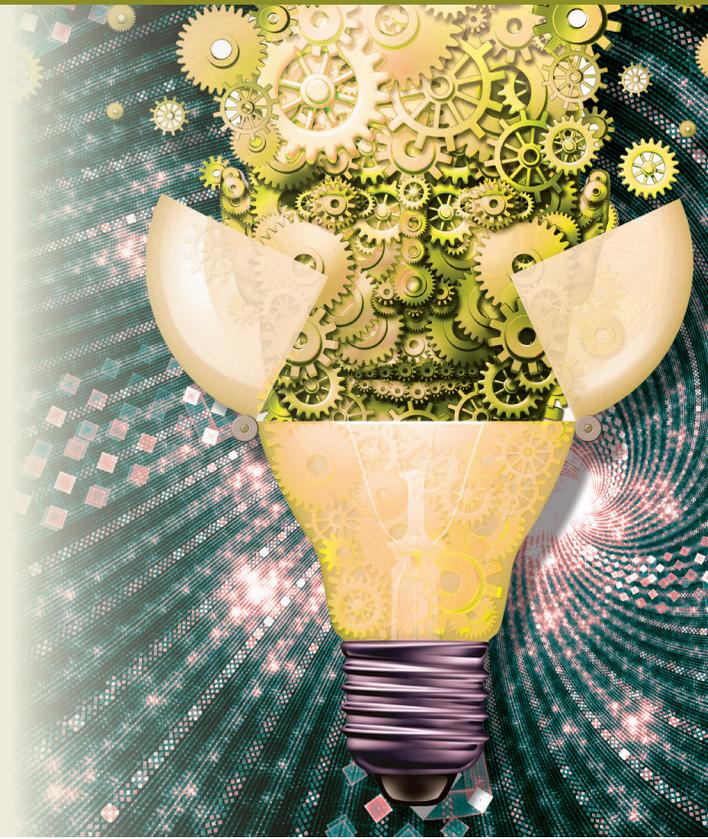


**database**  
TRENDS AND APPLICATIONS

2018 | AUG

# The Imminent Convergence of IoT and MACHINE LEARNING by Joe McKendrick

MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE ARE CHANGING THE VERY MEANING OF DATA ANALYTICS. SOME SAY THE RISE OF AI AND MACHINE LEARNING HERALDS A NEW INDUSTRIAL REVOLUTION.



There is widespread adoption of machine learning, though most organizations are still in the early stages with the technology. Machine learning, which can be broadly defined as “getting computers to act without being explicitly programmed,” has been adopted by 25% of enterprises, according to a recent survey conducted by Unisphere Research, a division of Information Today, Inc. Another 20% are considering adoption. At this time, a majority of companies with machine learning, 51%, have been using it less than 1 year.

The Internet of Things keeps rapidly expanding as well. Recent estimates from Juniper Research put IoT growth at 140% within the next 4 years, growing from 21 billion connected devices to 50 billion during that time. Growth will be driven by edge computing services (the processing of data away from the cloud and closer to the source), increasing the need for both deployment scalability and security.

One thing is certain: With its connected networks of vehicles, production facilities, machines, cameras, and sensors, IoT is generating enormous quantities of data. Time series inputs, energy usage, aircraft engines, and the testing of all of the above is generating gigabytes and petabytes of measurement data on a continuous basis. Identifying and

collecting all this data is one issue, and monitoring it is another. Ultimately, the ability to not only analyze the data but also deliver predictive and prescriptive insights from it, is the challenge organizations are now attempting to capitalize on.

As these two technology forces converge, the impact on organizations will extend well beyond the data center. Combined, they are emerging as prominent strategic weapons

*The ability to not only analyze IoT data but also deliver predictive and prescriptive insights from it, is the challenge organizations are now attempting to capitalize on.*

for forward-thinking digital enterprises. IoT, enhanced and made actionable, is a product of the digital age, in which data now flows from potentially every product, every customer touchpoint, and every corner of the

enterprise. The elevation of IoT-driven insights is becoming the backbone of today’s smart manufacturing initiatives.

These machine-learning-enhanced systems need to deliver the real-time data companies need to gain increased insights into the performance of the products they deliver, as well as act on opportunities for innovation and growth. Through this virtuous cycle, enterprises learn more and more about how their products and services are behaving, draw insights from that, and plug that understanding into new releases—and even back into their existing product base. Machine-learning-driven insights from IoT data adds intelligence to help adjust, update, and inform decision makers and enterprise-wide ecosystems.

To achieve this, machine learning has to be effectively implemented within IoT projects. However, the benefits of such efforts tend to be hazy, as the value of both IoT and machine learning is still being digested by business decision makers. The challenge—as is typical with sophisticated technology initiatives with many moving parts—is that implementation of IoT and machine learning is seen as a technology discussion, versus a business strategy.

A recent study, led by Mohammad Saied Mahdavejad of the Kno.e.sis—Ohio Center

## *Machine-learning-driven insights from IoT data adds intelligence to help adjust, update, and inform decision makers and enterprise-wide ecosystems.*

of Excellence in Knowledge-enabled Computing, concluded that IoT is clearly the leading source of new data flowing into enterprises, requiring that these enterprises step up their data science capabilities. “Since data is generated from different sources with specific data types, it is important to adopt or develop algorithms that can handle the data characteristics,” they point out. They also caution that the great number of resources that generate data in real time is not without the problem of scale and velocity. “Since IoT produces high volume, fast velocity, and varieties of data, preserving the data quality is a hard and challenging task. Although many solutions have been and are being introduced to solve these problems, none of them can handle all aspects of data characteristics in an accurate manner because of the distributed nature of big data management solutions and real-time processing platforms.”

In addition, the Kno.e.sis team observes, companies are still struggling to get a handle on IoT data itself. “The abstraction of IoT data is low, that is, the data that comes from different resources in IoT are mostly of raw data and not sufficient enough for

analysis,” they state, while also noting that solutions that have been proposed are incomplete. “For instance, semantic technologies tend to enhance the abstraction of IoT data through annotation algorithms, while they need more efforts to overcome its velocity and volume.”

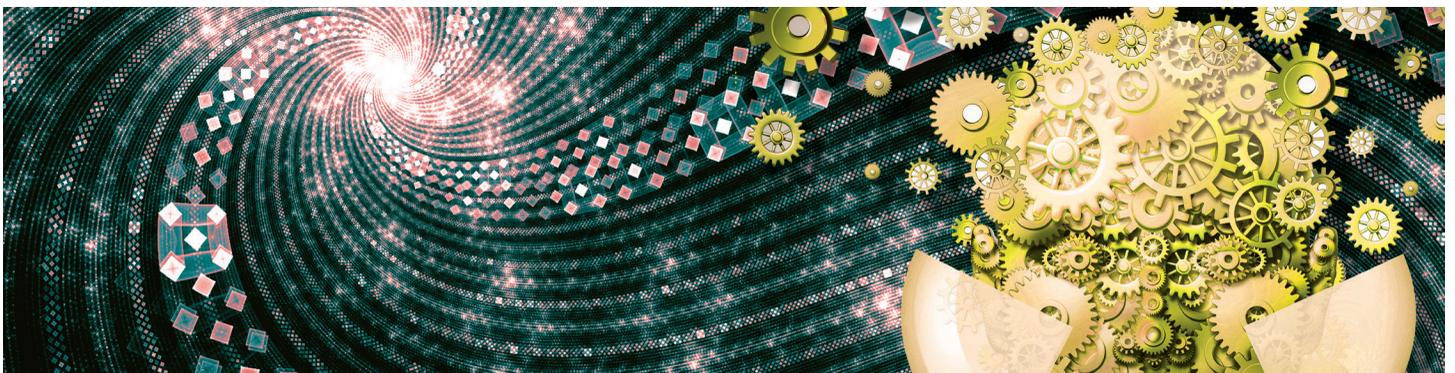
Ultimately, machine-learning algorithms will help overcome the issues with the big data associated with IoT, they point out, while advocating development of “smart data” as the optimal form of IoT-originated data. “IoT needs algorithms that can analyze the data which comes from a variety of sources in real time. For example, deep learning algorithms, evolutionized form of neural networks can reach to a high accuracy rate if they have enough data and time. Semi-supervised algorithms which model the small amount of labeled data with a large amount of unlabeled data can assist IoT data analytics as well.”

The ability to address these issues, and link IoT and machine learning into a platform that delivers real-time insights and responses, will deliver a range of benefits to enterprises. From a market-building

standpoint, it enables greater customer engagement: Building intelligence into products—through sensors and systems that track usage patterns and monitor wear and tear—can be a powerful service offering to engage customers. Whether it is downloadable software updates to vehicles, machines, or even consumer products such as speakers and thermometers, companies can also monetize such services, opening up new lines of business.

In addition, enterprises will benefit from reduced costs, especially internally, as machines, systems, and devices within the enterprise walls can be maintained or updated before the appearance of problems that result in production bottlenecks or expensive downtime. Even within data centers, the performance of systems, software, and devices can be monitored and addressed before IT administrators receive their wake-up calls at 4:40 on a Sunday morning.

It’s still the early days and companies have yet to get their arms around IoT and machine learning. As these two initiatives are joined, new directions for businesses will open up. ■





# Accelerate Enterprise Machine Learning With Cloudera

**MACHINE LEARNING HOLDS** the promise of augmenting human intelligence and automating decision making, transforming industries and creating new markets. Access to nearly unlimited data and computing power combined with rapid innovation in open source technologies are making machine learning capabilities broadly accessible. Yet adoption remains limited by key challenges related to data management, data science and machine learning workflow and productivity tools, and the ability to apply emerging machine learning technologies to solve the right business problems quickly, at scale.

Machine learning depends on data, and many organizations still struggle with foundational data management. Data needs to be collected, processed, stored, and organized for broad enterprise consumption spanning use cases, from data warehousing to data science to real-time applications. Yet departments often manage data and analytics in silos, where each group may have its own team, infrastructure, data, tooling, and process.

And while data science is a team sport, teams typically lack the productivity tools to work together at scale. Building machine learning models requires access to granular data, use of the latest open source tools, bursts of computing power, and tools to reproduce and share research. Thus, data scientists frequently copy data to their laptops or anywhere they can use their preferred tools, creating security risks and challenges for IT in keeping data secure and governed. And deploying models to production is even more challenging than exploratory data science, often requiring time-consuming and error-prone recoding and complex DevOps knowledge.

Finally, uncertainty in a rapidly changing technology landscape, obscurity of best practices for building and organizing teams, and difficulty separating hype from

reality make it difficult for organizations to define an actionable strategy and practical roadmap to applying and scaling machine learning for business. Without a focus on strategy and practical business problems, it's all too easy to embark on expensive "AI" science projects that consume time and resources but fail to deliver results.

These challenges are resulting in missed opportunities to take advantage of data, creating missed growth opportunities, increased costs, security and data privacy risks, and attrition of frustrated data professionals. Cloudera addresses these challenges with comprehensive solutions for machine learning in the enterprise, empowering organizations to own the future of their business.

For an ideal platform for data management, Cloudera's Enterprise Data Hub (EDH) with Shared Data Experience (SDX) unifies data collection, access, and deployment with shared security and governance. It enables data scientists to access unlimited data of all types—structured and unstructured—in a common repository that enables transparency and reproducibility. The EDH platform works on your choice of storage, on-premises or cloud, and supports common analytics disciplines including data science, data warehouse and BI, real-time streaming and IoT workloads, and ETL and data engineering. Unlike typical data centers and cloud environments where workloads operate in siloed, single-purpose application stacks, each with its own storage, security, governance, and schema, Cloudera Enterprise with SDX delivers a common services framework that includes a shared data catalog, security, governance, and lifecycle management. Providing consistent data context across the analytic functions in a unified platform makes it faster, easier, and safer to develop and deploy high-value data science and machine learning use-cases.

Cloudera Data Science Workbench (CDSW) balances the needs of both IT and data science teams, making secure, collaborative data science at scale a reality for the enterprise and accelerating the delivery of new data products. IT can deliver managed, secure, self-service access to corporate data through the compute power and distributed storage of the Cloudera platform. At the same time, data scientists have the freedom to use their favorite open source tools with full project isolation and reproducibility, as well as publish models as reports, batch jobs, or REST APIs. The shared environment enables reuse, further improving productivity.

Finally, accelerating time to results with machine learning requires more than software. Cloudera Fast Forward Labs offers applied machine learning strategic advisory services and research, providing clarity and practical guidance for busy organizations, increasing their productivity while saving them time and money. Cloudera Fast Forward Labs' dedicated research staff continually monitor the latest techniques and industry best practices, determining how best to apply them to difficult business problems. Their published research saves executives and data science teams time, so they can be focused on building the future of the business, and advisory services help customers understand where to focus valuable resources, how to implement effective practices, as well as expediting development and helping avoid technical dead ends.

Owning the future of your data-powered business requires the right data platform and tools combined with expert guidance for solving the right problems, quickly and at scale. ■

---

**CONTACT CLOUDERA TODAY**

to unlock the transformational potential of machine learning for your business.