Unleashing IT

BIG OPPORTUNITY

Strategies, solutions, and advice for leveraging big data and unlocking your competitive edge

Unleashing IT

UNLOCK YOUR COMPETITIVE EDGE

Big data presents tremendous opportunities for every company in every industry. It's purely a matter of finding the right use cases, the right strategies, and the right partners that can turn those opportunities into real business value.

This edition of *Unleashing IT* highlights the strategies, solutions, and providers that can help you unlock your competitive edge.

You'll get insight and updates from leading Hadoop vendors like MapR (page 6), Hortonworks (page 8), and Cloudera (page 14). You'll learn about solution bundles that take the guesswork and complexity out of big data deployments (page 5). And you'll hear how the global leader in web and video conferencing (page 16) and one of America's most populous cities (page 18) are transforming their data resources into business insights.

What's the common element behind their big data success? A rock solid, tightly integrated infrastructure that brings together Hadoop clusters, analytics engines, legacy systems, and both old and new data sources. One that helps turn big opportunity into big value.

For more information, follow the links inside or contact Cisco at 1-800-553-6387 and select option 1 to speak with a Cisco representative. We welcome your feedback on the articles in this publication at UnleashingIT.com.

Sincerely,

Jim McHugh
Vice President
Cisco Systems, Inc.

Shannon Poulin
Vice President
Intel Corporation

STRATEGIES & SOLUTIONS

EXPERIENCES

BIG DATA ON-RAMPS

The race is on. Industry experts explain the road ahead and how to start your big data engines.

After years of hype and development, big data has crossed the proverbial chasm from trendy concept to bona fide business enabler. According to Gartner[1], 73 percent of organizations have invested or plan to invest in big data in the next two years. But because it is still relatively new, many of these investments are focused on pilots, proofs of concept (POCs), and early deployment projects.

"Most companies are keenly interested in big data," says Dave Kloempken, global sales director for data center systems at Cisco, "but there's a lot of confusion about how to get started and how to show value. It's a process to find the right use cases, but once customers do, the deployments expand rapidly."

"Companies that get past the initial hurdles are seeing great results," adds Brandon Draeger, director of business development for big data solutions at Intel. "Research from Bain and Company has shown that companies successfully leveraging big data are not just more profitable, but also better able to understand and meet their customers' needs."

WHY BIG DATA? WHY NOW?

The explosion of computing devices and applications is at the heart of today's big data fervor. The growing volume, variety, and velocity of modern data sets—often called the "three Vs"—are not only a challenge for organizations, but also an incredible opportunity. Once captured, organized, and analyzed, big data can lead to differentiating insights that can be turned into business value.

But it's not that simple. The proliferation of unstructured data is now greatly outpacing the structured and refined information typically found in an enterprise data warehouse.

And these conventional database environments were never designed for the "three Vs," making them a poor fit for big data purposes.

"Traditional data warehouses can't accommodate unstructured data and can be extremely expensive to scale," says Kloempken. "That's why Hadoop has emerged so quickly."

"It's really about using the right tools for the job," says Draeger. "Ninety percent of new data being created is semi-structured or unstructured and is expected to reach a scale of 40 exabytes by 2020. Conservatively, that is about 5.2 terabytes for every person on earth."

THE RISE OF HADOOP

An open source software project that enables the distributed, fault tolerant processing of large data sets across clusters of servers, Hadoop is quickly rising to the big data challenge. Designed for unstructured data, Hadoop can merge conventional reports with a wide variety of data sets—from web logs and sensors to Facebook and Twitter—all in one database. It's also 20 to 50 times less expensive than traditional methods of data management and storage.

"Data warehouses aren't going away. They're still better suited for many important, repeatable needs, like operational reporting," says Kloempken. "But because of its cost, flexibility, and scalability, Hadoop is extremely attractive for a number of newer data sources and use cases."

These use cases don't always start with big data intent, says Bob Fosina, big data software and solutions specialist at Cisco. Many organizations are adopting Hadoop for cost saving purposes and then using the platform for big data experimentation.

"Companies are learning that they can use Hadoop to offload their ETL [extract, transform, load] processes, or pull out some of the stale data that is crowding their data warehouse, or reduce the cost of mission-critical software licensing," Fosina explains. "There are a number of compelling opportunities to reduce storage and licensing costs while at the same time creating a platform for big data POCs."

These "big data on-ramps," as Fosina likes to call them, are turning Hadoop into a mainstream data management and storage platform. And adoption will only increase as countless pilot projects blossom into full-blown production deployments.

STARTING SMALL

Whether Hadoop is deployed for immediate cost savings, long-term business transformation, or both, Kloempken recommends starting small with big data. Data warehouse optimization is often the first step, laying the foundation for operational analytics and business insights that grow over time.

"Once you have the platform, you can define a use case, pull in a few data sources, and conduct a pilot project," says Kloempken. "And that will invariably lead to other use cases. It tends to snowball."

And for good reason. According to the MIT Sloan Center for Digital Business[2], data-driven enterprises outperform industry peers by up to six percent and are up to 26 percent more profitable. The race is on, and the road ahead is more defined than ever before.

If you haven't already, it's time to start your big data engines.

[1] http://www.gartner.com/newsroom/id/2848718

[2] http://www.capgemini.com/resource-file-access/resource/pdf/The_Digital_Advantage__How_Digital_Leaders_Outperform_their_Peers_in_Every_Industry.pdf

To register for an upcoming Big Data Virtual Event series, visit UnleashingIT.com/virtualconference.

REFERENCE ARCHITECTURES AND SOLUTION BUNDLES NOW AVAILABLE

Hadoop clusters, analytics software, and supplementary big data technologies all have one thing in common: A fundamental need for a rock solid, tightly integrated infrastructure on which to operate.

"Big data is all about pulling a wide variety of data sources into a common environment," says Raghunath Nambiar, distinguished engineer and chief architect of big data solutions for Cisco. "That requires alignment and orchestration spanning compute, storage, networking, and application environments."

Cisco and its big data partners offer just that. Through the Intel® Xeon® processor-based Cisco Unified Computing System™ (Cisco UCS®) Integrated Infrastructure for Big Data, organizations have access to a set of reference architectures and solution bundles that are designed for big data deployments. Validated designs are available for leading Hadoop platforms and analytics software.

"The architectures can be used as is or customized to meet specific business requirements," says Nambiar. "We also have a number of single SKU bundles that make it easy to order and deploy big data environments."

The new Cisco UCS Director Express for Big Data enables end-to-end cluster management and monitoring, he adds. And Nexus 7000 and 9000 switches in tandem with Cisco Application Centric Infrastructure deliver best-in-class scalability.

"With these reference architectures, solution bundles, and UCS Director Express, organizations can focus squarely on innovation and business outcomes," says Nambiar. "They don't have to worry about piecing together the infrastructure or optimizing it for the Hadoop and analytics software they plan to use."

To learn more, see Raghunath Nambiar's blog at UnleashingIT.com/rn.

BENCHMARKING BIG DATA

The Transaction Processing Performance Council (TPC) launched TPCx-HS in 2014. The first—and to date, the only—vendor-neutral, industry standard big data benchmark, TPCx-HS measures the price, performance, and energy consumption of big data systems like Hadoop.

"Benchmarking provides an objective assessment of system performance and other variables," says Raghunath Nambiar, distinguished engineer and chief architect of big data solutions for Cisco. "This is not only important for comparing and selecting technology options, but also for driving faster, less expensive, and more energy efficient systems."

TPCx-HS is designed to measure both hardware and software, including Hadoop Runtime, Hadoop Filesystem API compatible systems, and MapReduce layers. It can also be used to assess a broad range of system topologies and implementation methodologies, in a technically rigorous and directly comparable, vendor-neutral manner.

Cisco and MapR recently announced the first TPCx-HS results, setting the high performance benchmark for big data systems.

"We published not one but three results," says Nambiar, "at 1TB, 3TB, and 10TB Scale Factors."

The benchmark configuration includes the Intel® Xeon® processor-based Cisco Unified Computing System™ (Cisco UCS®) Integrated Infrastructure for Big Data, two redundant active-active Cisco UCS 6296 Fabric Interconnects running Cisco UCS Manager version 2.2, 16 Cisco UCS C240 M3 Servers running Red Hat Enterprise Linux Server 6.4, and MapR Distribution with Apache Hadoop.

"TPC benchmarks are highly regarded in the industry and we are pleased to establish these high-performance standards for Hadoop," says Jon Posnik, vice president of business development at MapR. "These results validate the scale, performance, and availability of the MapR Distribution. We're already seeing tremendous demand for the combined solution with Cisco to enable utility-grade Hadoop environments."

To register for an upcoming Big Data Virtual Event series, visit UnleashingIT.com/virtualconference.

MAPR EXECUTIVE: ARCHITECTURE MATTERS

Big data success demands a data layer, application layer, and hardware platform that all work in concert.

From an architectural standpoint, big data is the biggest thing to hit the data center in decades, says Jack Norris, chief marketing officer for MapR, which offers an enterprise-grade Hadoop platform. And it pokes holes in traditional assumptions and methods.

"The separation of compute and storage, how data is handled and processed … all of those things are being reconsidered," Norris explains. "A lot of it is based on speed. Companies are figuring out how to do business as it happens."

This means making recommendations and offering special deals before a customer checks out, for example. Or identifying cyber security threats before sensitive data has been compromised. Or anticipating maintenance needs before equipment and heavy machinery go down.

DISTRIBUTED, FAULT TOLERANT PROCESSING

Conventional solutions and approaches—which typically require different data sets from different systems to be refined and normalized—make real-time decision making difficult, if not impossible. With new data platforms that support distributed, fault tolerant processing, however, organizations can reimagine how they gather, store, and take advantage of internal and external data sources.

"Companies no longer have the luxury of time to scrub and transform their data," says Norris. "They need an underlying, distributed data layer like Hadoop, which can handle traditional, mission-critical data as well as the massive amount of unstructured data being generated every second of every day."

But a flexible, scalable data layer isn't enough, he adds. It must be combined with an equally capable application layer and hardware platform that all work in concert.

"Architecture matters," Norris insists.

GET AN ENTERPRISE-GRADE HADOOP DEMO

MapR delivers on the promise of Hadoop with a proven, enterprise-grade platform that supports a broad set of mission-critical and real-time production uses. The company brings unprecedented dependability, ease-of-use, and world-record speed to Hadoop, NoSQL, database, and streaming applications in one unified distribution for Hadoop.

To learn more about Hadoop architectures and use cases for creating an as-it-happens business, sign up for a demo and live discussion at UnleashingIT.com/demo.

He points to the combination of MapR's Hadoop Distribution and the Cisco Unified Computing System™ (Cisco UCS®), powered by Intel® Xeon® processors, as an ideal architecture for big data. As the architecture scales to hundreds or even thousands of nodes, which is common in big data deployments, Cisco's Application Centric Infrastructure (Cisco ACI™) extends policy models across networks, servers, storage, security, and services.

"Get the architecture right, and everything else will be faster, easier, and more successful."

-Jack Norris, chief marketing officer, MapR

"You're only as good as your weakest link," says Norris. "You can't have latency in the network, and you can't have a shaky computing or storage foundation. With MapR, [Cisco] ACI, and [Cisco] UCS, organizations have a transformative big data platform that complements existing systems."

SPEEDING UP

As these new architectural paradigms become more commonplace, IT practitioners are being forced to evolve—and move faster.

"IT architects have been trained for years and years to slow down, understand every application, and define every database schema. Because if they got it wrong, they had to rearchitect everything. Even if they got it right, changes in the business required major updates," says Norris. "Hadoop is the opposite. There are no forced schemas, and there is a ton of flexibility. Things can be spun up in a day that would have previously taken weeks."

This requires new thinking, new skills, and new approaches. Norris suggests starting small by deploying a big data platform and experimenting with a defined use case. Offloading workloads from an enterprise data warehouse to reduce costs, pulling new data into existing applications for better customer service, and analyzing log files to improve information security are three common starting points.

"Deploy the platform, start to fill it with data, do a proof of concept, and expand over time," Norris recommends. "Whatever the use case, big data is incredibly transformational and will tangibly impact a business and its customers. The competitive advantage that it can deliver isn't hype, it's real. Hadoop users are proving this time and again."

And it all starts with the architecture.

"Get the architecture right, and everything else will be faster, easier, and more successful," says Norris.

MURKY WATERS: CLEARING THE CONFUSION SURROUNDING DATA LAKES

Industry insiders define and discuss the emergence of data lakes.

GET STARTED WITH ENTERPRISE HADOOP

Hortonworks Sandbox is a personal, portable Hadoop environment that comes with a dozen interactive Hadoop tutorials. It includes many of the most exciting developments from the latest HDP distribution, packaged in a virtual environment that you can get up and running in 15 minutes.

No data center, no cloud service, and no Internet connection are needed. To access Sandbox, visit UnleashingIT.com/hw.

Because of their ability to handle large volumes of incongruent data, Hadoop-based systems are becoming a foundational pillar of data storage and analytics. Unlike traditional relational databases that require information to be transformed into a specified structure, or schema, before it can even be loaded into the database, Hadoop focuses on storing data in its raw format. This allows analysts and developers to apply structure to suit the needs of their applications at the time they access the data.

"With the continued growth in scope and scale of applications using Hadoop and other data sources, the vision of an enterprise data lake has started to materialize," says Shaun Connolly, vice president of strategy for Hortonworks, a leading contributor to and provider of Hadoop for the enterprise. "Combining data from multiple silos, including internal and external data sources, helps organizations find answers to complex questions that no one previously knew how to ask."

DATA LAKE, DEFINED

The concept of a data lake is closely tied to Hadoop and its ecosystem of open source projects. The traditional "Schema On Write" approach of data management requires a lot of forethought and IT involvement, whereas Hadoop's "Schema on Read" approach empowers users to quickly store data in any format and apply structure in a very flexible and agile way, whenever needed.

As such, a data lake represents a shared repository where any type of data can be collected, accessed, and analyzed by any number of users within an organization.

But according to Gartner[1], the growing hype surrounding data lakes is causing substantial confusion in the information management space. Several vendors are marketing data lakes as an essential component of big data implementations, but there is little alignment between vendors about what comprises a data lake, or how to get value from it.

"In broad terms, data lakes are marketed as enterprisewide data management platforms for analyzing disparate sources of data in its native format," says Nick Heudecker, research director at Gartner. "The idea is simple: instead of placing data in a purpose-built data store, you move it into a data lake in its original format. This eliminates the upfront costs of data ingestion, like transformation. Once data is placed into the lake, it's available for analysis by everyone in the organization."

CONTROL VERSUS FREEDOM

While data lakes effectively solve a number of problems—overcoming independently managed data silos through consolidation and providing a new way to handle large quantities of disparate and often unstructured data—they also create gray areas surrounding information governance, control, and security.

"There is always value to be found in data," says Andrew White, vice president and distinguished analyst at Gartner. "But the question your organization has to address is this: Do we allow or even encourage one-off, independent analysis of information in silos or a data lake, bringing said data together, or do we formalize to a degree that effort, and try to sustain the value-generating skills we develop?"

According to Hortonworks, companies have an opportunity to do both.

"Hadoop complements existing systems and data warehouses, it doesn't replace them," says John Kriesa, vice president of strategic marketing at Hortonworks. "And it's the data lake that pulls it all together, from Hadoop and data warehouse environments to the systems, inputs, and sensors generating data to the analytics tools being applied."

For information and applications requiring the highest security and fastest, most consistent response, enterprise data warehouses are still the platform of choice, he adds, delivering the performance, governance, and control that is often needed for mission-critical activities. But when an organization wants to bring together and explore both structured and unstructured workloads from a variety of systems and silos, Hadoop environments and data lakes provide a compelling and cost-effective option.

"Implementing Hadoop is typically a journey from single instance applications to a fully-fledged data lake," says Connolly. "The journey is not about assembling petabytes of data. It's about encouraging people to combine new types of data with existing data sources, and enabling them to play with that data in ways that creatively unlock the value within."

[1]http://www.gartner.com/newsroom/id/2809117

EASING THE TRANSITION TO HADOOP

Virtualization and automation software can greatly reduce the time, complexity, and risk of offloading processing and storage from data warehouses to Hadoop.

Many companies are in the process of evaluating Hadoop—not just for big data analytics, but for a more cost-effective and scalable storage option than traditional data warehouses. But according to industry experts, the transition isn't always easy.

"People underestimate the effort and complexity of porting data to Hadoop," says Murthy Mathiprakasam, principal product marketing manager for Informatica, a leading provider of data integration software. "Hadoop is so new, not everyone understands how to use it effectively. And many try to move data manually, writing their own scripts and schema, which is time consuming and can be error prone."

To port data into Hadoop, it must first be collected and processed. And even after it has been moved, the data still requires a level of refinement. Get one of these steps wrong, and there can be undesired ripple effects. Bugs can emerge. There can be inconsistencies with the data. Or worse yet, business decisions can be made based on incomplete information and inaccurate assumptions.

FACILITATING GRADUAL TRANSITIONS

Fortunately, data virtualization and automation software can help ease Hadoop transitions.
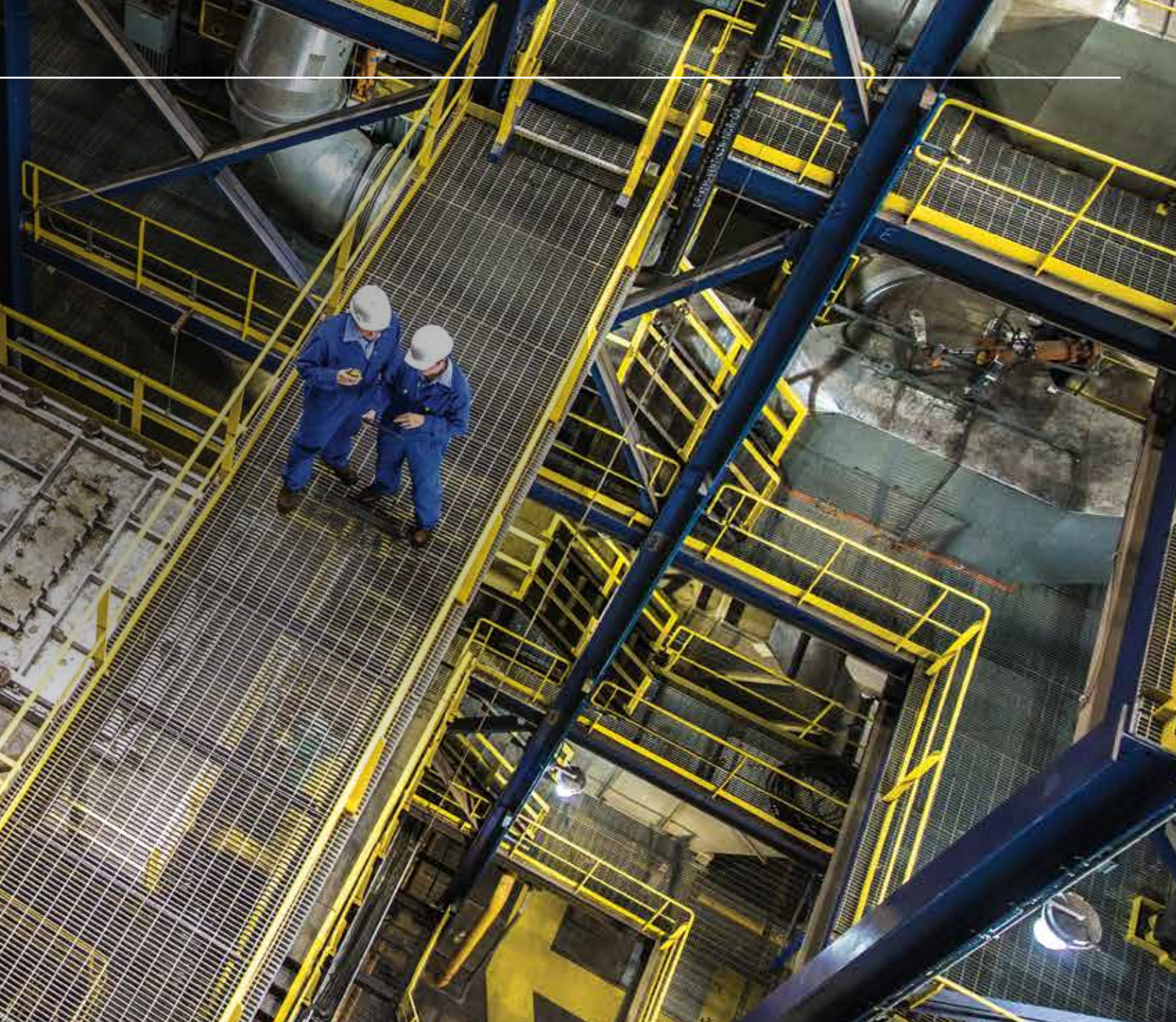
"These things don't happen overnight," Mathiprakasam explains. "Automation software reduces the time and risk of porting the data from one place to the next. And virtualization enables the data to remain accessible during the transition, no matter where it is or where it will reside."

He points to the combination of Cisco Data Virtualization, Informatica Big Data Edition, and the Intel® Xeon® processor-based Cisco Unified Computing System™ (Cisco UCS®) as an ideal platform for offloading processing and storage from data warehouses to Hadoop. Informatica Big Data Edition enables companies to run data transformation and data quality processes using a simple visual development environment

on Hadoop clusters installed on Cisco UCS servers. And the distributed data environments can be federated using Cisco Data Virtualization to provide business intelligence and analytics with a single point of access to all data.

"Hadoop is efficient and it's inexpensive," says Mathiprakasam. "But it isn't easy or simple. It's complex. The combination of Cisco Data Virtualization, Informatica software, and Cisco UCS can greatly reduce the time, effort, and risk of putting your data into a Hadoop cluster."

To learn more, register for Informatica events, webinars, white papers, and more at www.dwoptimization.me.

THE NETWORK: A TREASURE TROVE OF INFORMATION

Shedding light on valuable—but often "dark"—assets.

Amidst the scuttle to collect and mine data from new sources like sensors and social media, a treasure trove of information—the enterprise network—is sometimes overlooked. Leave it to Cisco to shed light on these valuable but often "dark" assets.

"People think of Cisco as a hardware company," says Scott Bergquist, senior director of Connected Analytics at Cisco. "But we've developed some powerful software and services that are based on 25 years of network innovation and leadership."

He's referring to Cisco's Connected Analytics™ solutions, which tie intelligence to infrastructure for new, network-based insights. Compared to other big data search and analytics engines that are broad in nature, Connected Analytics are ready made software solutions designed for very specific use cases.

"There is too much data moving too fast across a network environment to push it all to a centralized repository," Bergquist explains. "So we push analytics to the data to

derive valuable insights in real time, while also serving up metrics and data to downstream applications and engines. In essence, we bring clickstream-type analytics to the hardware platform."

Purchased independently or as a service, Connected Analytics packages are available for network deployment, collaboration, contact center, service providers, retail, events, mobility, and threat analytics. All of them provide analytics wherever they are needed—at the network edge in real time, or in the data center—for faster insights and decisions.

"If a company has a data lake or analytics engine, we can add data and value to that environment," says Bergquist. "And we can also solve very specific challenges and provide high-value intelligence, whether it's for a contact center or a retail environment."

For a Forrester Total Economic Impact study on Cisco Data Virtualization and Connected Analytics, visit the resource center at UnleashingIT.com.

▌ RECONSIDERING DATA VIRTUALIZATION

An increasingly important middleware layer, data virtualization can bring together big, complex data environments.

It's been said that data virtualization is less relevant with the explosion of big data platforms that can accommodate massive quantities and various types of information. Data experts at Cisco patently disagree.

"With enterprise data warehouses, Hadoop clusters, and cloud-based data sources, everything is still very siloed," says Peter Tran, product marketing manager at Cisco. "Pulling it all together into a single repository is unrealistic, but you still need access to all that data, no matter where it is."

That's where virtualization shines. Cisco Data Virtualization software augments conventional data integration by combining data sources to optimize query, compute, and network infrastructures. It allows companies to access and query all types of data across the network as if it is in a single place.

"It isn't a data store, but rather a middleware layer that delivers 'virtual' data to the right place," explains Rick Schreiber, director of data virtualization services at Cisco. "It's the layer between data sources—which can number in the hundreds or even thousands—and the applications that consume the data, like analytics engines."

Without virtualization, data is continually replicated as it is moved from one place to another, or as new schema are applied for analytical or reporting purposes. This not only results in data sprawl, but also unnecessary complexity and cost. With virtualization, however, data can be accessed and utilized without being physically moved. This means less formatting, less replication, faster response, and better scalability.

"Data virtualization is a perfect fit for big, complex data environments," says Tran. "It complements enterprise data warehouses and Hadoop clusters. In fact, it can help align those two environments, delivering data between them and pulling data from them for business-critical applications."

GET THE eBOOK

For an eBook and case study on Cisco Data Virtualization, visit the resource center at UnleashingIT.com.

e

NEW PROCESSORS DESIGNED FOR REAL-TIME ANALYTICS, HADOOP ENCRYPTION

Through product enhancements and collaborative efforts, Intel and Cloudera are accelerating big data innovation.

Two new processor families designed for data management, analytics, and security are already having an impact on big data innovation. The Intel® Xeon® processor E7 v3 product families deliver real-time analytics as well as enhanced performance and reliability for data-intensive computing. And the Intel Xeon processor E5 v3 accelerates Apache Hadoop HDFS encryption, enabling a 2.5X improvement in encryption off-load.

Both have spawned new technical advances while helping optimize core Hadoop technologies. Cloudera, a leader in enterprise analytic data management and an Intel partner, is utilizing the processors in four releases of its Hadoop distribution to improve data performance and protection. As a result, full database encryption is now possible with minimal impact to system performance, allowing processing resources to run Hadoop jobs at a faster rate.

"Partnering with Intel over the last year has allowed us to accelerate Hadoop innovation by committing to strengthen open source Hadoop projects in areas such as performance, security, and manageability," says Tom Reilly, chief executive officer at Cloudera. "It is our mutual goal to ensure big data is at its full potential and that our customers can deploy an enterprise data hub where and how they need to access existing assets and investments."

A FASTER PATH TO INSIGHTS, OUTCOMES

Together, Cloudera and Intel have created a faster path for enterprises across the financial services, telecommunications, healthcare, retail, and technology industries to better manage data, derive insights faster, and deliver transformative business outcomes. Numerous enterprises are already taking advantage of the new capabilities.

Caesars Entertainment, for example, is using the Cloudera distribution to improve the customer experience in its casinos and entertainment venues. The solution has helped reduce processing time for key jobs from six hours to 45 minutes and expand data analysis to include both unstructured and semi-structured data. A global leader in healthcare technology, Cerner Corporation has built a Cloudera-powered enterprise data hub to bring together a vast number of data sources and create a more complete picture of a person's health to predict potential risks with better efficacy. And dating site eHarmony is utilizing the Cloudera distribution to perform complex analyses that help facilitate more personalized results and improve the chances of relationship success.

"By bringing Intel's strengths in data center technologies to our joint efforts with Cloudera to enhance performance, scalability, and security in Apache Hadoop, we are delivering faster innovation in data management," says Diane Bryant, senior vice president and general manager of Intel's Data Center Group. "By uniting our technical efforts and fostering a robust partner ecosystem, we are accelerating the deployment of big data solutions, making it easier for companies of all sizes to obtain clear business value from data."

LEARN MORE

For self-guided, interactive Apache Hadoop demos and tutorials, access Cloudera Live at UnleashingIT.com/ah.

BECOMING A DATA-DRIVEN BUSINESS

Utilizing Platfora's Big Data Discovery Platform, Cisco WebEx is becoming more efficient and improving the quality of its products and services.

With 93 percent of Fortune 100 companies using its collaboration solutions, Cisco WebEx® is already the global leader in web and video conferencing. But there's always room for improvement.

"We want to be more operationally efficient, make better decisions, and deliver a better quality of service," says Joe Hsy, director of software engineering at Cisco. "It all starts with data."

And WebEx has a lot. Roughly 12 million WebEx meetings are conducted each month, with more than 50 million attendees from 230 countries on average. What represents some 20 billion minutes of collaboration every month also generates an avalanche of data that can be collected and analyzed. According to recent estimates, WebEx will collect petabytes of data over the next two years alone.

"We're making the transition to a true data-driven business," says Hsy.

To do so, the organization needed a new big data discovery platform. One that would provide better scalability at a lower cost than its pre-existing data warehouse, and simplified analytics for non-technical business users.

ANALYTICS FOR ALL

According to Hsy, WebEx's incumbent data warehouse was expensive to scale. It required an inordinate amount of time and effort to clean and prepare workloads before they could be analyzed. And it did not provide a clear path for business users to access and examine the data. Fortunately, WebEx was able to solve these challenges by leveraging Platfora's Big Data Discovery Platform and Cloudera's Hadoop cluster running on the Intel® Xeon® processor-based Cisco Unified Computing System™ (Cisco UCS®).

"Traditional tools can't handle the volume or variety of data that modern enterprises need to analyze to stay competitive," says David Littlewood, senior director of alliances at Platfora. "Analytics need to be easy to use for business people, and they need to be fast. The combination of Platfora, Hadoop, and Cisco UCS provides a complete big data platform that removes complexity and rapidly delivers business insights."

With the new platform in place, WebEx was able to define three preliminary use cases focused on meeting join time, survey data, and fraud analytics. These use cases were not driven forward by WebEx technical staff, but rather product management, marketing, and operations teams.

FAST INSIGHTS

Meeting join time is a critical performance indicator for WebEx, showing how quickly attendees are able to get into a meeting and directly impacting their overall experience with the service. This seemingly simple metric is made up from a variety of sources, including web, server, gateway, user, and back-end system data.

"With Platfora, the team was not only able to bring all those data points together to provide a coherent view of meeting join time, they were able to view the process in a completely new way," Hsy explains. "Just knowing where in the process a delay might occur doesn't tell you much. You need to know why it is happening there. Is it the time of day? The number of callers in the meeting? The device that is being used?"

Based on the analytics, WebEx was able to pinpoint certain regions, Internet service providers, and customers that were experiencing meeting delays. With many of those issues involving traffic routing and security scans, WebEx was able to proactively communicate—and in some cases, resolve—the underlying sources of latency.

WebEx is also using the new data platform to draw parallels between customer surveys and product performance.

"We had previously built some analytics tools that focused primarily on audio and visual quality, but we could never correlate and analyze both product and business performance," says Hsy. "With our new platform, we can analyze end-to-end processes and get a better feel for the full customer experience."

They can also detect patterns and anomalies, such as illicit use of the service. In pulling together a variety of data sources, WebEx is not only able to identify these patterns, but also the who, where, and when of fraudulent activities.

SHARING DATA AND ANALYTICS

WebEx is now expanding its use of analytics and pursuing a number of new use cases. For example, the organization is examining user telemetry to better understand customer preferences and the most and least used features.

Those insights will be utilized to optimize products and pricing, and to develop a recommendation engine that improves upselling.

WebEx also plans to share its analytical capabilities with others. A new portal is in the works that will allow WebEx customers and partners to access and analyze data about their usage of the service.

"With self-service capabilities built-in, we can segment and expose data in a secure and confidential way," says Hsy, "allowing our business users, our customers, and our partners to conduct their own analytics and gain new insights."

SCHEDULE A BIG DATA DISCOVERY WORKSHOP

Platfora's Big Data Analytics platform is the fastest way to go from raw data to analytics. Born in the big data era, Platfora's end-to-end platform replaces the need for ETL, data warehousing, and business intelligence tools.

To schedule a complimentary discovery workshop about the ways Hadoop and Platfora can be leveraged in your organization, contact David Littlewood at davidl@platfora.com.

PROMINENT CITIES TAPPING BIG DATA

With the help of World Wide Technology, one of the largest cities in the United States is identifying big data opportunities and formulating a strategy for success.

Leading web-based companies, well-known retailers, financial titans, and healthcare conglomerates aren't the only ones taking advantage of big data. Some of the world's most prominent cities are also learning how to exploit their information resources in new and impactful ways. And World Wide Technology (WWT), a global technology integrator with $6.7 billion in annual revenue, is helping them.

"When it comes to big data, some start with technology and figure it out from there," says Chris Infanti, engagement manager for WWT's big data practice. "We take the opposite approach, focusing squarely on business outcomes."

Defining the opportunity is the first step, he explains, and that's exactly what WWT has done for one of the most populous cities in the United States. Instead of diving deeply into IT circumstances and desires, WWT reached out to 11 of the city's departments, including police, fire and rescue, public works, and human resources.

"The city's central IT group wasn't used to collaborating with the various departments," Infanti says. "Starting with the business groups allowed for more ideas and free thinking."

WWT conducted nearly 30 two-hour interviews with city employees, seeking to understand how each department serves the city and its citizens, what applications they utilize, their goals and challenges, and how data can potentially help. The discovery process took three weeks and included more than 100 city employees.

"We found that the city is already capturing loads of data," says Infanti. "But much of that data is fragmented and dispersed among 300 departmental applications."

Therein lies the big data opportunity. WWT used the findings to identify 32 use cases, most of which involve combining disparate data sets for greater efficiency and value. WWT also developed an overarching data management strategy and roadmap for the city.

PRIORITIZING OPPORTUNITIES

With the city's data assessment and initial planning completed, WWT has recommended prioritizing the use cases and selecting two or three that offer ease of implementation and high impact. The city can then deploy a big data platform that is sized for those use cases and capable of scaling over time.

"It's always good to show the value of an investment quickly," Infanti says. "And once you successfully complete the first use case, others become easier."

Some of the use cases are very specific, such as combining property data with police records to improve the safety and efficiency of building inspections. Others are broader, such as visualizing the full scope and status of projects that involve multiple city departments. And some take advantage of unstructured data, like video from sanitation trucks and police dashboard cameras, text information from the city's 3-1-1 system, and building plans and survey data used by

the public works department. The city is also considering predictive models for preventive maintenance on crucial city infrastructure.

"With 32 use cases that can all deliver value, it's a matter of prioritization," says Infanti. "Fortunately, with new platforms like Hadoop, the city can start small and expand over time."

Before any capital investments are made, the city can use the WWT Advanced Technology Center (ATC) in St. Louis to test different software solutions on the Intel® Xeon® processor-based Cisco Unified Computing System™ and see how a big data platform would integrate with existing systems. And because big data technologies are so new for the city and its staff, the ATC can also be used for hands-on education and training.

In addition to pinpointing opportunities to leverage data in new ways, the discovery and planning process has also brought the city's IT group closer to its departments. And for the first time, they are working together to envision and design a centralized data infrastructure, policy, and roadmap that will help the city work more efficiently in support of its citizens.

SCHEDULE A WWT BIG DATA WORKSHOP

World Wide Technology uses a proven and innovative approach to help large public and private organizations explore, evaluate, architect, and implement advanced technology.

To register for a complimentary two- to four-hour technical and strategic whiteboard session on big data opportunities and planning, email BigData@wwt.com or visit www2.wwt.com/request-workshop.

THE INTERNET OF EVERYTHING

presents

the last checkout line

We're building the Internet of Everything for business. With UCS Server solutions providing data-center performance everywhere, mobile applications and analytics deliver, and customers pay as they shop. Let's confine the checkout line to yesterday.
See how at cisco.com/tomorrowstartshere

CISCO ™

TOMORROW starts here.

Cisco UCS with Intel® Xeon® processors

intel inside™ XEON®