



Solution Brief

Big Data in the Cloud: Converging Technologies

How to Create Competitive Advantage Using Cloud-Based
Big Data Analytics

Why You Should Read This Document

This paper describes how cloud and big data technologies are converging to offer a cost-effective delivery model for cloud-based big data analytics. It also includes:

- How cloud computing is an enabler for advanced analytics with big data
- How IT can assume leadership for cloud-based big data analytics in the enterprise by becoming a broker of cloud services
- Analytics-as-a-service (AaaS) models for cloud-based big data analytics
- Practical next steps to get you started on your cloud-based big data analytics initiative

Contents

- 3 The Cloud as an Enabler for Big Data Analytics
- 6 Cloud and Big Data: A Compelling Combination
- 9 IT as a Broker of Cloud Services
- 11 Next Steps for IT

The Cloud as an Enabler for Big Data Analytics

Two IT initiatives are currently top of mind for organizations across the globe: big data analytics and cloud computing. Big data analytics offers the promise of providing valuable insights that can create competitive advantage, spark new innovations, and drive increased revenues. As a delivery model for IT services, cloud computing has the potential to enhance business agility and productivity while enabling greater efficiencies and reducing costs.

Both technologies continue to evolve. Organizations are moving beyond questions of what and how to store big data to addressing how to derive meaningful analytics that respond to real business needs. As cloud computing continues to mature, a growing number of enterprises are building efficient and agile cloud environments, and cloud providers continue to expand service offerings.

It makes sense, then, that IT organizations should look to cloud computing as the structure to support their big data projects. Big data environments require clusters of servers to support the tools that process the large volumes, high velocity, and varied formats of big data. Clouds are already deployed on pools of server, storage, and networking resources and can scale up or down as needed. Cloud computing offers a cost-effective way to support big data technologies and the advanced analytics applications that can drive business value.

The paper will describe:

- How cloud computing is an enabler for advanced analytics with big data
- How IT can assume leadership for cloud-based big data analytics in the enterprise by becoming a broker of cloud services
- Analytics-as-a-service models for cloud-based big data analytics
- Practical next steps to get you started on your cloud-based big data analytics initiative

Big Data Trends

What makes cloud computing such a cost-effective delivery model for big data analytics? How are big data and cloud technologies converging to make big data analytics in clouds a reasonable option? For big data analytics:

Data is becoming more valuable. Today the conversation is shifting from “What data should we store?” to “What can we do with the data?” Enterprises are looking to unlock data’s hidden potential and deliver competitive advantage. Gartner predicts that enterprise data will grow by 800 percent from 2011 to 2015, with 80 percent unstructured (for example, e-mails, documents, video, images, and social media content) and 20 percent structured (for example, credit card transactions and contact information).¹

What Is Big Data Analytics?

Big data refers to huge data sets that are orders of magnitude larger (*volume*); more diverse, including structured, semistructured, and unstructured data (*variety*); and arriving faster (*velocity*) than you or your organization has had to deal with before. This flood of data is generated by connected devices—from PCs and smart phones to sensors such as RFID readers and traffic cams. Plus, it's heterogeneous and comes in many formats, including text, document, image, video, and more.

The real value of big data is in the insights it produces when analyzed—discovered patterns, derived meaning, indicators for decisions, and ultimately the ability to respond to the world with greater intelligence. Big data analytics is a set of advanced technologies designed to work with large volumes of heterogeneous data. It uses sophisticated quantitative methods such as machine learning, neural networks, robotics, computational mathematics, and artificial intelligence to explore the data and to discover interrelationships and patterns.

With the potential for so much data to reveal insights that can boost competitiveness, companies must find new approaches to processing, managing, and analyzing their data—whether it's structured data typically found in traditional relational database management systems (RDBMSs) or more varied, unstructured formats. Plus, combining diverse data sources and types has the potential to uncover some of the most interesting unexplored patterns and relationships.

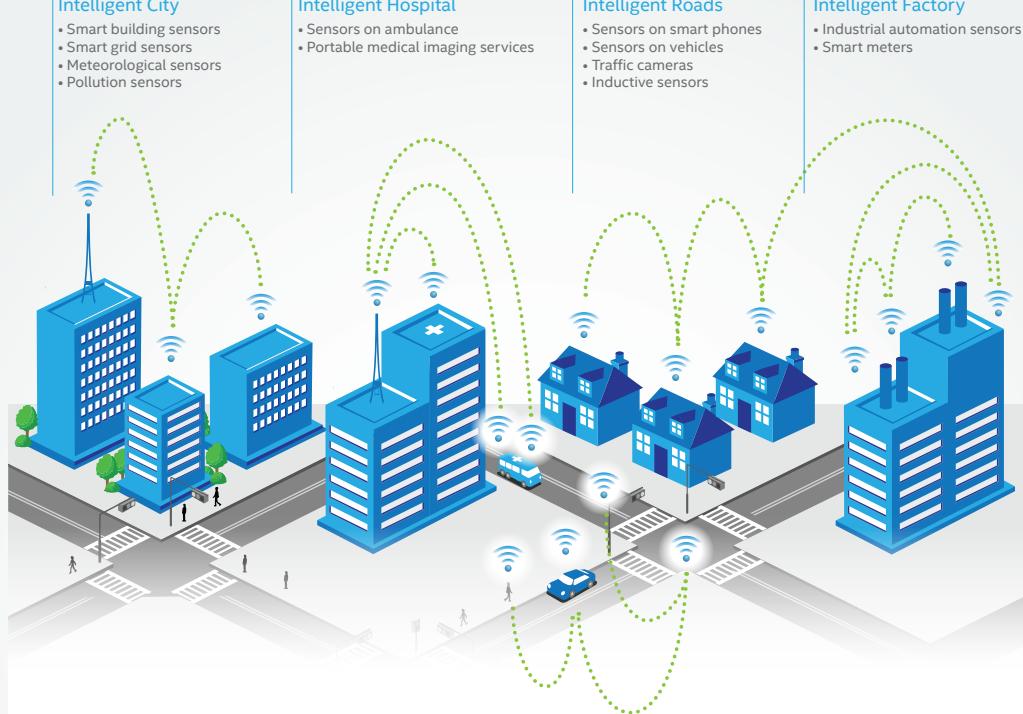
Data analytics is moving from batch to real time. Intel's 2012 survey of 200 IT managers in large enterprises found that while the amount of batch versus real-time processing is split evenly today, the trend is toward increasing real time to two-thirds of total data management by 2015.² At the same time, the technology for processing real-time or near-real-time information is moving past hype to early stages of maturity.

Real time supports predictive analytics. Predictive analytics enables organizations to move to a future-oriented view of what's ahead and offers organizations some of the most exciting opportunities for driving value from big data.

Real-time data provides the prospect for fast, accurate, and flexible predictive analytics that quickly adapt to changing business conditions. The faster you analyze your data, the more timely the results, and the greater its predictive value.

The scope of big data analytics continues to expand.

Early interest in big data analytics focused primarily on business and social data sources, such as e-mail, videos, tweets, Facebook* posts, reviews, and Web behavior. The scope of interest in big data analytics is growing to include data from intelligent systems, such as in-vehicle infotainment, kiosks, smart meters, and many others, and device sensors at the edge of networks—some of the largest-volume, fastest-streaming, and most complex big data. Ubiquitous connectivity and the growth of sensors and intelligent systems have opened up a whole new storehouse of valuable information. Interest in applying big data analytics to data from sensors and intelligent systems continues to increase as businesses seek to gain faster, richer insight more cost-effectively than in the past, enhance machine-based decision making, and personalize customer experiences.



Big Data in Context: Smart City Example

In addition to the transactional, social, and location data generated by people, device sensors generate in real time some of the fastest-growing big data. Processing and analytics can be applied to these valuable data sources via provisioned embedded, cloud, or dedicated IT infrastructure and storage and high-performance computing solutions.

Cloud Technologies Mature

Cloud computing is becoming a reality for many businesses, with private cloud deployments often leading the way. Cloud technology is maturing and addressing barriers to adoption with improvements in security and data integration, while IT organizations are evolving to support cloud services delivery. As a result, businesses are demonstrating growing trust in cloud delivery models. For example, a 2013 survey from Ubuntu found that 55 percent consider the cloud ready for mission-critical workloads.³

Organizations continue to store more and more data in cloud environments, which represent an immense, valuable source of information to mine. Plus, clouds offer business users scalable resources on demand. Combining the Intel® Xeon® processor-based servers and storage, along with Intel SSDs and Intel 10 GbE networking resources used in cloud environments, with big data processing tools like Apache Hadoop* software provides the high-performance compute power needed to analyze vast amounts of data efficiently and cost-effectively. Running Hadoop* in virtualized environments continues to evolve and mature with initiatives like VMware's open-source project Serengeti*, among others.

Cloud and Big Data: A Compelling Combination

Cloud delivery models offer exceptional flexibility, enabling IT to evaluate the best approach to each business user's request. For example, organizations that already support an internal private cloud environment can add big data analytics to their in-house offerings, use a cloud services provider, or build a hybrid cloud that protects certain sensitive data in a private cloud, but takes advantage of valuable external data sources and applications provided in public clouds.

Using cloud infrastructure to analyze big data makes sense because:

Investments in big data analysis can be significant and drive a need for efficient, cost-effective infrastructure.

The resources to support distributed computing models in-house typically reside in large and midsize data centers. Private clouds can offer a more efficient, cost-effective model to implement analysis of big data in-house, while augmenting internal resources with public cloud services. This hybrid cloud option enables companies to use on-demand storage

space and computing power via public cloud services for certain analytics initiatives (for example, short-term projects), and provide added capacity and scale as needed.

Big data may mix internal and external sources. While enterprises often keep their most sensitive data in-house, huge volumes of big data (owned by the organization or generated by third-party and public providers) may be located externally—some of it already in a cloud environment. Moving relevant data sources behind your firewall can be a significant commitment of resources. Analyzing the data where it resides—either in internal or public cloud data centers or in edge systems and client devices—often makes more sense.

Data services are needed to extract value from big data.

Depending on requirements and the usage scenario, the best use of your IT budget may be to focus on analytics as a service (AaaS)—supported by your internal private cloud, a public cloud, or a hybrid model.

Unlocking the Potential of Big Data in Clouds

Cloud computing models can help accelerate the potential for scalable analytics solutions. Clouds offer flexibility and efficiencies for accessing data, delivering insights, and driving value. However, cloud-based big data analytics is not a one-size-fits-all solution.

Organizations using cloud infrastructure to provide AaaS have multiple options. By weighing factors of workload, cost, security, and data interoperability, IT can choose to utilize their private cloud to mitigate risk and maintain control; use public cloud infrastructure, platform, or analytics services to further enhance scalability; or implement a hybrid model that combines private and public cloud resources and services.

The bottom line: No matter which cloud delivery model makes the most sense, businesses with varying needs and budgets can unlock the potential of big data in cloud environments.

Analytics as a Service Insight Framework

You can address user needs across the full range of analytics requirements with cloud-based AaaS—from data delivery and management to data usage. By developing a comprehensive cloud-based big data strategy, you can define an insight framework and optimize the total value of enterprise data.

An AaaS insight framework encompasses the following key capabilities:

- **Capturing and extracting structured and unstructured data from trusted sources**, including prioritizing the most critical data and identifying what to retain and for how long
- **Managing and controlling data under comprehensive policy and governance guidelines** across a global enterprise and in compliance with specific industry requirements
- **Performing data integration, analysis, transformation, and visualization** to deliver the right information to the right location at the right time

Cloud Service Types for AaaS

AaaS can be deployed in the cloud based on various cloud service types. Determining the right mix of services depends on user needs weighed against existing internal resources—such as a private cloud environment—that are already in place.

The basic cloud service types for analytics as a service include infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

Infrastructure as a Service (IaaS)

Deployed on-premise or via a cloud provider, IaaS enables you to allocate or buy time on shared server resources, which are often virtualized, to handle the computing and storage needs for big data analytics. Cloud operating systems manage high-performance servers, network, and storage resources.

IaaS provides the foundation for many companies' cloud services. However, IaaS also requires greater investment of IT resources in the context of implementing big data analytics. Your organization will be responsible for installing your own software, such as the Hadoop framework, or a NoSQL database, such as Apache Cassandra*, MongoDB*, or Couchbase* technologies. Your team will also be responsible for managing your assigned resources—which can be made easier with automated tools for management and resource orchestration.

Platform as a Service (PaaS)

PaaS provides developers with tools and libraries to build, test, deploy, and run applications on cloud infrastructure. PaaS reduces management workload by eliminating the need to configure and scale elements of your Hadoop implementation and serves as a development platform for advanced analytics applications.

Software as a Service (SaaS)

Specific applications for cloud-based big data analytics can be provisioned with SaaS. You may need to use multiple SaaS applications to cover the range of scenarios business users require. For example, software that works well for sentiment analysis may not work for risk management or asset performance. SaaS can be offered as a standalone application or part of a greater cloud provider solution. For example, Karmasphere offers a pay-as-you-go application that analyzes data stored with Amazon* S3 using Amazon Elastic MapReduce.

Infrastructure-as-a-Service (IaaS) Examples

The following is a sample of IaaS solutions from providers in the cloud technology ecosystem.

- Amazon* Web Services
- Citrix* CloudPlatform
- Windows Azure* and Microsoft* System Center
- OpenStack* software
- Rackspace*
- Savvis*
- Verizon* Terremark*
- VMware vCloud* Suite

Platform-as-a-Service (PaaS) Examples

The following is a sample of PaaS solutions from providers in the cloud technology ecosystem.

- Force.com
- Google* App Engine
- Red Hat* OpenShift*
- VMware Cloud Foundry
- Windows Azure*

Software-as-a-Service (SaaS) Examples

The following is a sample of SaaS solutions from providers in the cloud technology ecosystem.

- Amazon* Elastic MapReduce
- Cetas* by VMWare* analytics solutions
- Google* BigQuery services
- Rackspace* Hadoop* service
- Windows Azure* HDInsight*

Intel Infrastructure Technologies for Cloud and Big Data Analytics

Servers based on the Intel® Xeon® processor E5 and E7 families are at the heart of infrastructure that supports both cloud and big data environments providing industry-leading, highly efficient, high-performance computing. In addition:

- Intel Xeon processor E5 family-based storage servers support advanced storage capabilities such as compression, encryption, automated tiering of data, data deduplication, erasure coding, and thin provisioning and are ideal for storing and processing large volumes of data. These compute-intensive storage tasks provide enhanced security, greater efficiencies, and better total cost of ownership through a reduced storage footprint.

- Intel Solid-State Drives (SSDs) are high-throughput, high-endurance drives for raw storage.
- Intel Ethernet 10 Gigabit Converged Network Adapters provide high-throughput connections for large data sets.

Intel also provides hardware-enhanced security capabilities, including Intel Data Protection Technology with Advanced Encryption Standard New Instructions (Intel AES-NI),⁴ which speeds data encryption and decryption up to 10 times.^{5,6} Also, Intel Platform Protection Technology with Trusted Execution Technology (Intel TXT)⁷ can provide a hardware root of trust to ensure that data is processed on or migrated to trusted pools of servers.

IT as a Broker of Cloud Services

Cloud computing and the myriad of public cloud services available to businesses have made it easier for two of the heaviest users of analytics—line-of-business owners and the chief marketing officer—to bypass IT and purchase services directly. However, uninformed business users may be tempted to buy “instant analytics,” but ad hoc methods for adopting public cloud services throughout your organization can cause significant problems—such as choosing the wrong vendor, losing control of your sensitive data, and getting a poor return on investment, to name a few. IT offers specific services, perspectives, and skills that can reduce the risk of using public clouds and better utilize existing private cloud resources.

Big data also demands a new set of skills in the enterprise—many of which reside in IT. IT departments can offer the technology know-how needed to help make cloud computing and big data work in your organization, including Hadoop administrators and developers and specialists in Hadoop components such as the Apache HBase* database. Big data analytics projects involve multidisciplinary teams, and IT members must be active collaborators with data scientists, another emerging big data-related role. Data scientists are individuals who apply big data to complex business problems and make sense of the results. While they may sit in the business, they also can be part of the IT organization.

As the broker for cloud services, IT can work with business users to get the best cloud-based analytics solution possible by making sure these important areas are considered.

Area	Questions
Institutional data management	What cloud providers are being evaluated? The potential for data to be stored, managed, and analyzed by multiple providers with no oversight is a huge risk.
Data ownership	Who owns the data that your provider stores and manages? Does your company retain ownership?
Security	What level of security is provided, and at what levels of the solution stack? Providers with security built deep into infrastructure and platform as well as at the application level can provide greater assurance.
Compliance	Are compliance issues addressed in general, as well as related to your specific industry? How is data anonymized to protect privacy?
Data integration	How is data integrated, and at what cost?
Data migration	How much data needs to be moved, and at what cost? Moving large volumes of data to and from the cloud may be cost prohibitive.
Data streaming	Is the data source streaming real-time information? Real-time data requires enormous resources to manage, and data that streams nonstop may be better handled in-house.
Technology evaluation	What big data storage, processing, and analytics solutions are provided? How are the components optimized for performance?
Skills requirements	What skills are needed to identify appropriate data sources, apply the right statistical and analytics models, and interpret results? Does the service provider include access to technology and analytics support? At what level? How does that complement in-house skills?
ROI	What is the return on investment for the technology, delivery model, security, and data integration methods?

IT Playing to Win with Big Data Analytics

In a growing number of companies, business users already consume IT as a service. IT can continue to extend this role to brokering cloud-based big data analytics services. As a cloud services broker, your role is to weigh user needs against the available delivery options for your organization. This means developing a strategy for private, public, and hybrid services; driving discipline into the selection of cloud service providers; and negotiating and establishing contracts with potential cloud service providers, among other similar tasks. Organizationally, this can reduce risk and better utilize existing investments in private cloud technologies. Individual users benefit by getting the right solution to meet their needs.

IT can quickly demonstrate value to the business by partnering with users to:

- Select the right private or public cloud implementation for their needs by defining technology requirements, assessing risk, and specifying deployment requirements based on corporate governance policies and regulatory compliance requirements. For example, certain workloads may have to be managed in a private cloud in a specific location.
- Build or work effectively with a technology partner to develop services as required.
- Evaluate and vet outside services for design, delivery, customization, pricing, privacy, integration, security, and support.

- Provision services from internal and external sources so that they appear seamless to users.
- Develop relationships with vetted cloud service providers.
- Manage existing services, including service level agreements (SLAs) and service life cycle.

As a service broker, IT collaborates with the business on the best way to use technology for competitive advantage. With cloud-based big data analytics, the objective must be to provide the right solution for users' needs balanced against corporate governance policies, existing IT resources, performance requirements, and overall business goals.

In most IT departments today, providing this consultative approach to service will require IT to reorganize to remove silos, hire or develop team members with new skills, and encourage a strong partnership with the business.

The payoff will be significant, especially for big data analytics projects, which require collaboration between IT technology experts, business users, data scientists, and others who can help develop the appropriate analytics plan and algorithms to extract meaningful insights from the data.

Next Steps for IT

IT is in a unique position in the organization. Despite—or maybe because of—explosive data growth, emerging technologies, and rapid change, you can provide much-needed leadership within your organization for big data analytics. First and foremost, consider how IT can evolve as the broker for big data analytics cloud-based services for your business. You can also:

- **Partner with business owners now** to help determine how big data can be used to solve your organization's business problems and align on opportunities. As a fully engaged partner, you can help evaluate and influence the choice of technology and establish best practices.
- **Explore technology options for cloud-based big data analytics**, including private, public, and hybrid delivery models. Keep up to date with trends, watch the market, and understand costs.

- **Create or update an existing big data strategy** that defines the process for engaging IT for big data analytics projects. Keep in mind that you will have to make it easy and fast for users to move forward, or business units will take matters into their own hands.
- **Consider how to organize IT** to better engage with business users and collaborate and consult on big data projects.

To learn more about big data and cloud computing, take advantage of resources available at the Intel IT Center. Visit intel.com/bigdata and intel.com/cloudcomputing.

Endnotes

1. Groenfeldt, Tom. "Big Data—Big Money Says It Is a Paradigm Buster." *Forbes* (January 6, 2012). forbes.com/sites/tomgroenfeldt/2012/01/06/big-data-big-money-says-it-is-a-paradigm-buster/
2. Peer Research: *Big Data Analytics: Intel's IT Manager Survey on How Organizations Are Using Big Data*. Intel IT Center (August 2012).
3. *Ubuntu 2013 Server and Cloud Survey*. Ubuntu Server (September 10, 2013).
4. No computer system can provide absolute security. Requires an enabled Intel processor and software optimized for use of the technology. Consult your system manufacturer and/or software vendor for more information.
5. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests such as SYSmark* and MobileMark* are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
6. Source: Testing with Oracle* Database Enterprise Edition 11.2.0.2 with Transparent Data Encryption (TDE) AES-256 shows as much as a 10x speedup when inserting 1 million rows 30 times into an empty table on the Intel Xeon processor X5680 (3.33 GHz, 36 MB RAM) using Intel Integrated Performance Primitives (IPP) routines, compared with the Intel Xeon processor X5560 (2.93 GHz, 36 MB RAM) without Intel IPP.
7. No computer system can provide absolute security. Requires an enabled Intel processor, enabled chipset, firmware, and software, and may require a subscription with a capable service provider (may not be available in all countries). Intel assumes no liability for lost or stolen data and/or systems or any other damages resulting thereof. Consult your Service Provider for availability and functionality. Consult your system manufacturer and/or software vendor for more information.

Share with Colleagues



Legal

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION, OR SAMPLE. Intel disclaims all liability, including liability for infringement of any property rights, relating to use of this information. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Copyright © 2015 Intel Corporation. All rights reserved. Intel, the Intel logo, the Experience What's Inside logo, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

