Cognitive Computing and Big Data Analytics
By Judith Hurwitz, Marcia Kaufman and Adrian Bowles

CHAPTER

7

The Role of Cloud and Distributed Computing in Cognitive Computing

The ability to leverage highly distributed and cost-effective computing services has not only transformed the way software is managed and delivered but also has become the linchpin for commercializing cognitive computing. Large cognitive computing systems require a converged computing environment that supports a variety of types of hardware, software services, and networking elements that have to be workload balanced. Therefore, cloud computing and a distributed architecture are the foundational models required to make large-scale cognitive computing operational. This chapter provides an overview of distributed computing architectures and cloud computing models.

Leveraging Distributed Computing for Shared Resources

The cognitive computing environment must provide a platform that consolidates a massive amount of information from disparate sources and process that information in a sophisticated manner. The system must also implement advanced analytics to gain insights into complex data. Clearly, a single integrated system would be impractical because of the need to bring so many different elements together. This is where highly distributed environments supporting

cloud computing become the delivery platform of choice. The *cloud* is a method of providing a set of shared computing resources including applications, compute services, storage capabilities, networking, software development, variable deployment modalities, and business processes. Cloud computing allows developers to combine distributed computing systems into a set of shared resources that can be used to support large cognitive workloads. To achieve this goal, it is important to base cloud services on standards and standardized interfaces. These interfaces are defined by standards organizations providing a consistent specification that can be widely adopted by cloud providers. This chapter provides insights into the role of distributed cloud services in making cognitive computing a reality.

Consumers of cloud services, including firms building cognitive computing applications, benefit from the shared resource model, which enables them to pay by usage on systems that operate close to peak efficiency. Owning these resources requires an ongoing fixed cost for carrying excess capacity for anticipated peak loads. Having the ability to use these services on demand makes them affordable to a wide range and size of organizations.

Why Cloud Services Are Fundamental to Cognitive Computing Systems

A cognitive system requires the capability to leverage data sources and complex algorithms. The most efficient and effective means to operationalize cognitive systems is cloud computing because by design, they are built on distributed computing models. Without distributing computing capabilities via the Internet, the World Wide Web (or just "web") would never have existed. In fact, the web was designed to enable researchers to share documents, images, videos, or audio files by simply assigning addresses without regard to the meaning of the content. With cognitive computing the environment is optimized to support a massive amount of data that must be analyzed and organized based on patterns. For example, the source data may be distributed across hundreds of different structured and unstructured information sources. In order to orchestrate the access to these sources, the cloud environment may have a catalog, index, or registry of pointers to the data as well as metadata associated with key resources. There may also be a requirement for analytics that leverage high-powered computing capabilities. An additional benefit of using cloud computing and its underlying distributed model is the ability to access high-powered computing engines to solve complex science, engineering, and business problems on demand. Again, the organization would not have to purchase this high-end system, but rather can consume the computational services only when needed.

Characteristics of Cloud Computing

Although this chapter covers a number of models of cloud computing, some characteristics are common to all models. These include elasticity and self-service provisioning, metering of service usage and performance, and workload management. In addition, supporting distributed compute capabilities is instrumental to the cloud. All these services are required because of the dynamic nature of the cloud. The cloud is purpose-built so that it can support a number of different workloads and characteristics of those workloads. This section includes a discussion of these capabilities and characteristics.

Elasticity and Self-service Provisioning

Elasticity of a cloud service offers the ability for consumers to increase or decrease the amount of compute, storage, or networking they need to complete a task. Although the notion of adding services is available in other modes of computing, within a cloud environment, elasticity is intended to be an automated service that is controlled by a self-service function. This is especially important when the consumer of a cloud service needs to increase the amount of compute services, for example, when applying an algorithm to a complex set of data. When that calculation is complete, the amount of compute resources can be automatically decreased. Within elasticity are the areas of scaling and distributed processing.

Scaling

With cloud elasticity you can scale the service to process shifting workloads. The two primary models of scaling are horizontal and vertical. *Horizontal scaling* (often called scaling out or scaling in) means that the same type of service is expanded based on the need of the workload. As more of the same capability is needed, the system allocates more resources. As the need diminishes, those resources are released to the pool. With horizontal scaling, additional servers or blades can be added to support expanding requirements. In contrast, *vertical scaling* (often called scaling up) occurs when one computing resource is expanded, creating a better match between the workload and the computing environment. Rather than adding more servers, a scale-up environment enables you to add additional memory or storage to the existing system environment. Vertical scaling is useful for solving problems with applications requiring highly distributed computing environments. As an example, Hadoop is designed to distribute computation across nodes, so it benefits from scaling up.

Distributed Processing

With the growth in big data, the ability to distribute processing across compute nodes to gain better performance is increasingly important. Although the idea

of a distributed filesystem is not new, new data technologies like NoSQL, HBase, and Hadoop are driving the importance of this capability. Using clusters of machines within a cloud to process complex algorithms is critical. Cognitive computing requires not only ingesting data but also the ability to analyze complex data to provide potential answers to complex problems.

Cloud Computing Models

Although there is sophisticated technology within cloud computing, you need to understand that cloud computing is a service-provisioning model that is transforming the degree to which companies can access and manage complex technology. The economic benefits of the cloud model are obvious. By providing a shared services model, each user pays only for services used. Within cloud computing models, there are a number of different approaches that are optimized for executing specific tasks for specific workloads. This is analogous to how power grids operate. A large metropolitan area does not have a single power plant to support all customers. Rather there is a system or grid that coordinates a highly decentralized set of power distribution stations supporting different neighborhoods. A well-designed power grid would model the distribution of power based on environmental conditions, consumption patterns, or catastrophic events. Having a shared services model makes a power grid affordable. If you take this analogy a step further, companies and individuals that invest in solar panels connected to a grid may get paid commensurate with the amount of electricity they contribute back to the public power grid. Likewise, in systems such as the World Community Grid, individuals contribute computing resources to help solve major computational problems. In the future, there may be cognitive computing grids where resources are shared across companies, industries, regions, and nations.

There isn't a single model of cloud computing. Rather there are a number of deployment models, including public, private, managed services, and hybrid clouds. Each of these deployment models features such service models as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Following is a description of each of these deployment models. The next section provides an overview of the technical underpinnings of each of these models (see Figure 7-1).

The Public Cloud

The *public cloud* is a utility model of computing typically offered as a shared multitenant environment, in which multiple users physically share a container within a single server. A *multitenant cloud* is a publicly accessible service that is owned and operated by a third-party service provider and is accessed through

an Internet connection. The customer pays based on usage or per unit of computing or storage. Therefore, a public cloud is often thought of as a commodity service. Typically, a user gains access to a service through a virtualized image—a combination of computing resources that can run independently from the physical hardware. Public clouds are most efficient when they support common workloads so that the system is automated and optimized for that workload. This is different than a data center where there may be multiple operating systems, and multiple types of applications and workloads. As such, it is difficult to optimize the environment for small numbers of simple workloads. The public cloud can be an effective economic model because it is built on a shared services model. The more customers that public cloud vendors such as Amazon.com, Microsoft Azure, and Google cloud services support, the less they charge per unit of usage. The typical payment model is based on a few cents per megabyte of storage or a unit of compute.

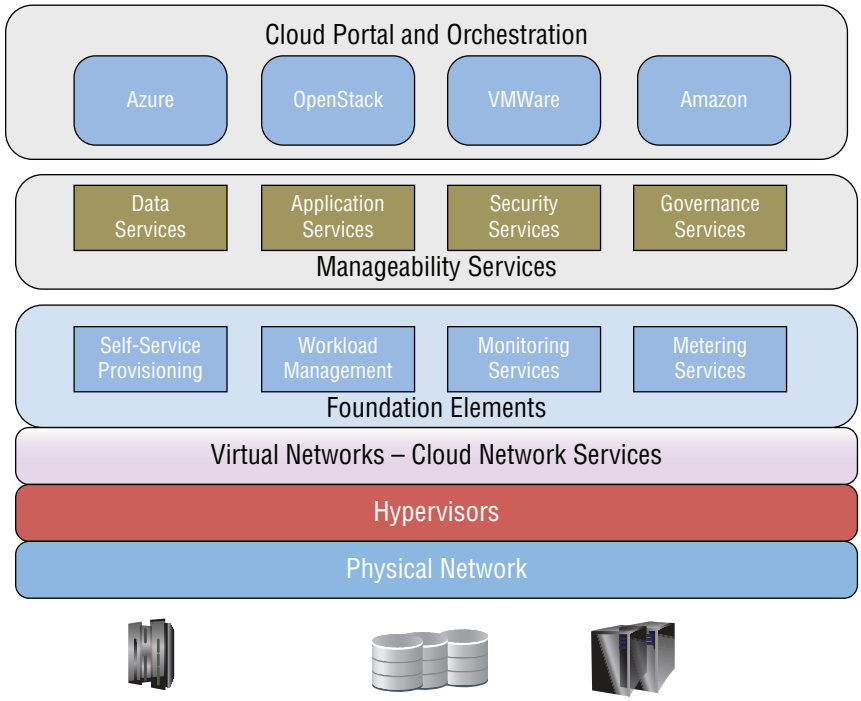

Figure 7-1: Foundations of a cloud architecture

One of the key characteristics of a public cloud is that it provides the same level of service and security to all its customers, represented by generalized service level agreements (SLAs). The service provider, therefore, manages its servers, automation, and security as an integrated environment. The customer leveraging these public services has little or no visibility into the resources within the operation of the system.

Well-designed commercial public cloud services tend to offer a reasonable level of service and security. Companies that need to have compliance and governance guarantees because of government regulations may be unable to use a public cloud service. In those cases, private cloud services may be a more viable option for critical customer and financial data. However, other commodity services such as e-mail are often implemented using public cloud services because they are not strategic assets for organizations. Some public cloud service providers offer additional specialized services such as virtual private networks or specialized governance services.

The Private Cloud

As the name implies, a *private cloud* is managed within a company's data center, and those resources are typically not shared with other companies. Like a public cloud, the private cloud is intended to be an optimized environment so that it supports a single underlying operating system with an optimized set of management and automation services. Like public clouds, the private cloud also provides optimization of workloads to improve manageability and performance. Because a private cloud is controlled internally, it can optimize security based on industry governance requirements. In addition, the private cloud can be established with a specific level of service required to support customers and partners. The company has the capability to implement tools and services to monitor and optimize both security and service levels.

Managed Service Providers

In addition to private clouds that are owned and operated directly by a company, there are managed service providers (MSPs) that provide dedicated cloud services designed and managed by a third party for the benefit of a specific customer.

Companies that are uncomfortable with public cloud services may not want to operate their own private clouds. In addition, some companies want to leverage a sophisticated set of services that are not resident within their own environment. *Managed Service Providers* (*MSPs*) typically offer industry-specific cloud services available as an ongoing supported service or as an on demand service. These cloud services have characteristics of either a public or a well-architected private cloud in that they offer management, security, and automation. They may offer some services as a multitenant environment but also provide the option of providing customers with their own private hardware environment that is secured just for their use. A managed service provider may provide a cognitive computing service that is specific to one industry such as a technique for analyzing customer churn in a retail environment through the use of a machine-learning algorithm. Therefore, the MSP is a form of private cloud because a single customer can use it on a dedicated, physically partitioned basis.

However, like a public cloud, it can serve multiple customers from a common, albeit physically partitioned, infrastructure.

The Hybrid Cloud Model

A *hybrid cloud* offers the ability to either integrate or connect to services across public, private, and managed services. In essence, a hybrid cloud becomes a virtual computing environment that may combine virtualized services in a public cloud with services from a private cloud, a managed service vendor, and a data center. For example, a single company may use its data center to manage customer transactions. Those transactions are then connected with a public cloud where the company has created a web-based front end and a mobile interface to allow customers to buy products online. The same company uses a third-party managed service that checks credit for anyone trying to use a credit card to pay for a service. There may also be a series of public cloud-based applications that control customer service details. In addition, the company uses extra compute capabilities from a public cloud provider during peak holiday periods to make sure that the website does not crash when the system becomes overloaded.

Although each of these elements are all designed and operated by individual vendors, they can be managed as a single system. A hybrid cloud can be highly effective because as a distributed system, it can enable companies to leverage a series of services that are the best fit for the task at hand, as shown in Figure 7-2.

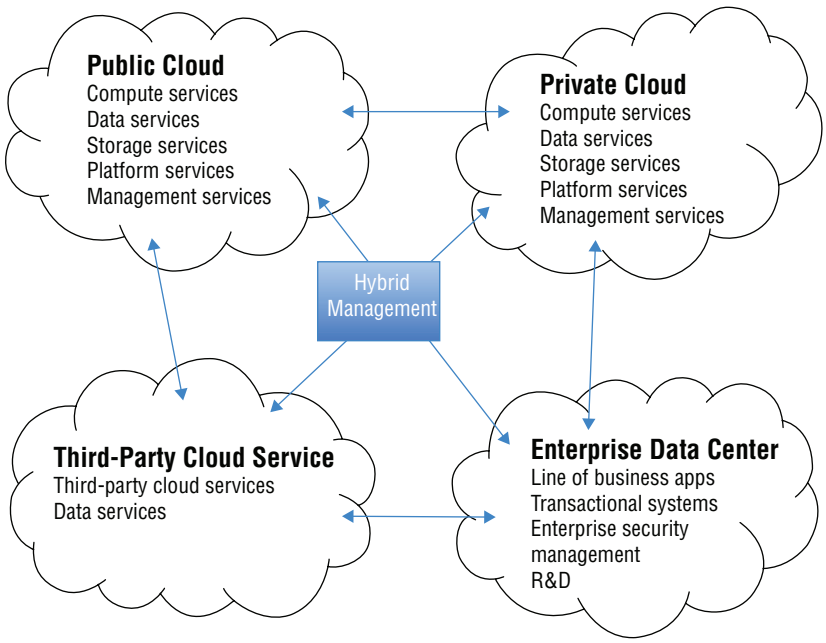

Figure 7-2: Hybrid cloud architecture

Whether the cloud model is public or private, they can allocate different cognitive-computing workloads to the optimal service components. Because the performance of a cognitive system can benefit from deployment on a variety of workload-optimized services rather than one unified system, a hybrid cloud model is the most logical and practical approach as these systems grow in production. Cognitive computing can be architected as a series of services that interact and call different services via Application Programming Interfaces (APIs) to execute a process or compute an algorithm in an efficient and cost-effective manner. When companies use big data sets to calculate complex algorithms on an occasional basis, it is often difficult to gain access to enough compute resources at a reasonable cost. For example, in the pharmaceutical industry, drug discovery requires the analysis of massive amounts of data that needs to be both stored and processed. Before cloud computing, these companies had to compromise and select only subsets of data to analyze. They had to be confident that the data they were selecting was the right subset. It was quite possible, however, that the patterns or anomalies might not have appeared in the subset or snapshot of data that they could afford to collect.

APPLYING CLOUD COMPUTING TO CLINICAL RESEARCH

One of the best ways to understand the benefits of cloud computing for sophisticated research is to look at a clinical research example, such as epilepsy analysis. In the December 10, 2013 issue of the *Journal of the American Medical Informatics Association*, the researchers discuss how they conducted epilepsy research by leveraging cloud-based big data analytics.

Researchers have been conducting research to attempt to discover new treatments for epilepsy (one of the most common neurological disorders). The typical source of data has come from recordings from electroencephalograms (EEG). This data is used to diagnose and evaluate epilepsy patients. If the signals from this data can be analyzed and visualized in real time, researchers can better determine what is happening with a patient before, during, and after an episode. In addition, this data can be correlated with ontologies designed to support conclusions between events and diagnosis.

Researchers working on this project found that if they could move from an integrated application that resided on a desktop to a cloud-based data management system, they could collect more data and analyze that data in real time. The researchers developed the Prevention and Risk Identification of SUDEP Mortality (PRISM) project. This web-based electrophysiology data visualization and analysis platform was called Cloudwave. This public cloud infrastructure integrated a patient information identification system and added a query system. The foundation of the system included the use of parallelized algorithms for computing using the MapReduce framework to interpret the huge volume of data. Data visualization correlates the results with the ontologies and other research such as databases of other risk factors. A query function can make the results accessible to researchers.

Without the support of cloud services and advanced analytics, the researchers would not have analyzed this amount of data in a reasonable timeframe. It would have taken too long and required the purchase of expensive hardware that was not in the budget of the organization. Even more important was that the real-time services were needed only intermittently and therefore a cloud service that could be used on an occasional basis was optimal for the requirements of the project. The organization also found a web service vendor that would support the Health Insurance Portability and Accountability Act (HIPAA) standards.

Delivery Models of the Cloud

Whether discussing public or private cloud deployment models, a number of important service delivery models define the way consumers and suppliers take advantage of these approaches to computing (refer to Figure 7-2). These models are divided into four different areas because they each provide a different capability that is important to implementing sophisticated services.

Infrastructure as a Service

As its name implies, *Infrastructure as a Service* (*IaaS*) is the foundational cloud service. IaaS provisions compute, storage, and networking services through either a virtualized image or directly on the computer systems. This is called native (or bare metal) implementation. Although bare metal implementations are frequently used when speed is the most important factor, the typical IaaS model relies on virtualization. A public IaaS service is designed as a self-service environment so that a customer can purchase a service such as compute or storage based on the instance of computing that is needed. Consumers can purchase an instance based on the amount of resources consumed over a specified period of time. When a consumer stops paying for the service, the resource disappears. In a private IaaS environment controlled directly by a company, those provisioned resources would remain in place and will be controlled by the information technology organization.

Virtualization

Virtualization is the technique that separates resources and services from the underlying physical delivery environment. In a traditional model, the hardware is partitioned through the use of a hypervisor. The *hypervisor* is software that provides a thin layer of code on top of the server that enables system resources to be shared. This means that a single system can support multiple operating systems, infrastructure software, storage, networks, and applications. In addition, the hypervisor enables more services to be supported on the same physical

infrastructure. IaaS relies on images that encapsulate the key capabilities required by a consumer to operate a cloud service such as an amount of computing capability or a set amount of storage. The image will include the capability to manage these resources such as add new code or balance the set of resources.

Software-defined Environment

The goal of IaaS is to optimize the use of system resources so that they can support workloads and applications with the maximum efficiency. A *Software Defined Environment* (*SDE*) is an abstraction layer that unifies the components of virtualization in IaaS so that the components can be managed in a unified fashion. In effect, the SDE is intended to provide an overall orchestration and management environment for the variety of resources used within an IaaS environment. Therefore, an SDE brings together compute, storage, and networking to create a more efficient hybrid cloud environment. It also enables developers to use a variety of types of virtualization within the same environment without the burden of hand-coding the linkages between these services.

Containers

A *container* consists of an application that is designed to run within IaaS, encapsulated together with its dependencies as a lightweight package ready for deployment. It includes well-defined and standardized Application Programming Interfaces (APIs) to make integration easier. A container is often used within the context of a Software Defined Environment. The use of containers creates an alternative to relying on virtualized images. Unlike virtualization, a container does not require a hypervisor. Several open source projects (e.g., Docker) have emerged in the past few years to facilitate this style of computing.

Software as a Service

Software as a Service (*SaaS*) is a defined application that is operated on a public cloud service. Today, virtually every enterprise software offering is available as SaaS, and it is becoming the de facto approach to desktop applications and personal software, as well. In fact, it is becoming difficult to buy or license some types of software because the SaaS model provides a more predictable revenue stream for the vendors.

SaaS applications are built to take advantage of IaaS. Therefore, like IaaS, SaaS is typically delivered in a multitenancy environment offering load balancing and self-service provisioning. This means that multiple users share a physical computing environment with other users and companies. Their own implementation is partitioned from other users. One of the benefits of SaaS is that the consumer is not responsible for software updates and maintenance of

the application. However, unlike a traditional on-premises application, the user does not have a perpetual license for the application. Rather the user pays on a per-user, per month, or per year basis. Many SaaS applications are designed as packaged applications based on a business process such as customer relationship management or accounting. These applications are designed in a modular fashion so that customers can select only what they require. For example, some accounting SaaS applications may have a foundation of a bookkeeping process and can expand into a complex online accounting system. Over the years, more and more areas of software are available as a service, including collaboration, project management, marketing, social media services, risk management, and commerce solutions.

SaaS implementations are expanding beyond the traditional packaged software. Increasingly, most emerging software platforms are implemented as cloud services as the preferred deployment model. One of the most important examples of the power of the cloud for cognitive computing is the advent of big data environments Hadoop and MapReduce that depend on a highly distributed cloud platform to process massive amounts of data. The distributed nature of the cloud enables complex computation to be completed quickly.

Business intelligence (BI) services have been available as cloud services for a number of years. However, the objective of these systems is to provide management with reports that capture historical performance of the business. Advanced analytics offerings are increasingly offered as cloud services. The complexity and amount of data analyzed demands the type of scalable and distributed properties of the cloud. The complex cognitive algorithms used in machine learning and predictive analytics are better served by a cloud infrastructure. One of the benefits of using the cloud for advanced analytics as a service is that it is more affordable for solving complex problems. For example, an analyst might need to build a predictive model to solve a specific problem in a quick timeframe. Rather than purchasing all the hardware and software, the analyst can leverage a sophisticated analytics application in the cloud. The analyst pays only for the capability used for that project. After the project is complete, there is no further financial obligation. The cloud offers the ability to solve a problem that leverages huge amounts of computing capability. There may also be the need to store the data and results from this analysis.

Analytics as a service in the cloud enables business managers or business analysts to leverage an analytics portal that documents best practices. Increasingly, there are offerings on the market that provide the knowledge of the data scientist without the expense of hiring those expensive resources. Many of these offerings enable a data scientist to optimize an algorithm based on the problem being solved. Some of the emerging use cases for analytics in the cloud come from industries such as retail, where managers want to understand what is driving profits in various business units. While it may be a simple question to ask, the answer is extremely complicated. The analysis requires the ingestion

of considerable information from a number of internal and external sources, followed by the computation to determine patterns, followed by recommendations for next best actions for issue resolution. A cloud analytic service can codify the best algorithm to apply to a specific analytical goal and the cloud service can access a specific capability on demand. In the future, analytics as a service will generate new models where the analytics service provider will provide services to help analyze a customer's data. In the long run, analytics as a service will enable data providers to provide new cognitive computing offerings. By the end of this decade, the major cognitive computing technologies such as Natural Language Processing (NLP), hypothesis generation/evaluation, and question answering systems should be available as standalone services, offered as SaaS components that can be integrated into a customer's application on demand.

Platform as a Service

Platform as a Service (*PaaS*) is an entire infrastructure package that is used to design, implement, and deploy applications and services in either a public or private cloud. PaaS provides an underlying level of middleware services that abstract the complexity away from the developer. In addition, the PaaS environment provides a set of integrated software development tools. In some cases, it is possible to integrate third-party tools into the platform. A well-designed PaaS consists of a orchestrated platform to support the life cycle of both developing and deploying software within the cloud. A PaaS platform is designed to build, manage, and run applications in the cloud.

Unlike traditional software development and deployment environments, the software elements are designed to work together through Application Programming Interfaces (APIs) that support a variety of programming languages and tools. Within the PaaS environment are a set of prebuilt services such as source code management, deployment of workloads, security services, and various database services.

Managing Workloads

The ability to manage workloads is at the heart of cloud computing. What makes cloud computing so powerful is that it enables an organization to bring together applications that live in the data center with those that reside on public and private clouds. To be operationally effective these various workloads have to act as a single, unified environment. In other words, these services need to be orchestrated together in a consistent manner. One of the fundamental approaches used to achieve this consistency is having the workloads abstracted from the underlying hardware environment.

Workload management in a traditional data center environment has been centrally controlled through job scheduling programs that orchestrated workloads in a serial and scheduled manner. The cloud environment is a different dynamic entirely because workloads are rarely scheduled in a predictable manner. Therefore, cloud workload management depends on load balancing—the process is designed so that complete workloads or components of a workload can be distributed across multiple servers within the cloud.

In a hybrid cloud environment the ability to manage overall performance requires monitoring of the overall service level of the servers, software, storage, and network. Any system, whether on premise or in the cloud, must be managed to achieve the contractual service levels required by customers. However, a cloud environment is more dynamic than an on-premise environment. Therefore, the system has to monitor performance and anticipate changes in compute requirements, the amount of data managed, or the addition of new workloads. A cognitive computing environment requires this type of flexible workload management because there is the requirement for sophisticated analytics of workloads. Data is constantly being evaluated and expanded as new sources of data become available.

Security and Governance

As cognitive solutions become a strategic platform for businesses, the ability to secure content and results becomes more important. No company will trust a system that may hold the potential for strategic differentiation if the information can be compromised. Therefore, security has to be defined at every level of the environment. Given the nature of the data within a cognitive computing system, it is critical that security is instituted so that unauthorized persons cannot access key data. Therefore, identity management will be key. You will need to work with the cloud provider to indicate which individuals with which roles are entitled to access or change data.

Any cloud environment will require the same levels of security as a traditional data center, including issues ranging from physical security of servers, storage, networks, applications, and data. In addition, there needs to be specific techniques for handling incidents, security of the specific applications, encryption, and key management.

Within a data rich environment, it is critical to be aware of the governance requirements to protect sensitive data. Different industries, markets, and countries have specific requirements for how data about individuals needs to be secured. For example, in the United States there is a regulation called The Health Insurance Portability and Accountability Act of 1996 (HIPAA) that requires that an individual's health information must be kept private. Countries such as Germany and France have specific regulations about where a person's data

can be stored. Therefore, although the cloud will have a set of data protections built into the environment, your company is still responsible for the protection of sensitive data. This is further complicated in a hybrid environment in which data may be distributed across a number of different public and private clouds.

Overall governance of data requires a strategy based on understanding the regulations of your industry and understanding how these regulations are implemented and executed in the various cloud applications and services used. Each company will have its own requirements to audit its own security, including its use of public and private clouds. Therefore, every organization needs to have a governance body in place that understands the cloud services used and how those companies comply with regulations. It is prudent, therefore, to create an overall governance plan incorporating every IT service used by your organization.

Data Integration and Management in the Cloud

Data integration in the cloud offers both huge potential and huge complexities. As with on-premise applications, most organizations have hundreds of different data sources that need to be managed. Although the availability of data in the cloud is a huge help in gaining access to critical information, it also means that there is a need to provide connections and techniques for integrating data sources. Simply connecting data does not solve the problem. Integrating data sources in the cloud requires the ability to correlate the relationships between sources through a catalog that defines the meaning of fields or data sources.

All data integrations are not created equal due to different requirements for each use case. For example, there are situations in which cloud data sources need to be tightly linked together because the sources are interdependent. This can be accomplished through data replication. In some cases, it is important to move several data sources into the same cloud environment for speed. In other situations, the original data source needs to remain in either a cloud data repository or within a data center. In this situation there is the need to provide pointers to move between sources. This typically happens when each source is independent. In fact, in most situations data will increasingly be managed in a distributed manner to process a large number of information sources that need to interact with each other.

Summary

Cloud computing is a critical deployment and delivery model for applications and data. The capability to distribute huge amounts of data is critical to the development of a cognitive system because it depends on the availability of the

right data sources that may physically reside in a hybrid environment. A cognitive system requires the capability to link to and manage the right data sources where they live, when they are needed. The cloud and distributed computing is one of the fundamental models to make it possible for a variety of data sources to be used in this level of decision making.