IBM Systems and Technology

Thought Leadership White Paper

June 2011

Building the perfect cloud with IBM System z

A cloud computing platform that leverages the strengths of industry-leading mainframe technology

Introduction

Cloud computing helps organizations transform business and technology. By freeing them from the responsibility of having to internally own, manage and maintain IT resources, cloud computing can help accelerate application development, rapidly provision test environments, speed time to market, and reduce development and test costs, among other benefits.

IBM® System z® technology offers the security, availability and virtualization capabilities that are critical in cloud computing. As a result, organizations that rely on System z are perfectly positioned to take advantage of the cloud. System z qualities of service provide unparalleled virtualization of resources, very high levels of utilization, flexible operations, massively scalable workloads and legendary security that is built into the hardware, operating system and subsystems.

Organizations have long known that System z is the ideal platform for databases. IBM leads the way in delivering reliable and secure mainframe database technology, including IBM DB2® for z/OS® and IBM Information Management System (IMS™). And most IT organizations are well on the way to virtualization with a variety of architectures. By using the highly available data stores of DB2 and IMS, which are easily hosted and managed by System z, organizations can leverage the capabilities of IBM zEnterprise™ System to manage not only the legacy System z environment, but also the Linux on System z environment.

zEnterprise is IBM's "system of systems," integrating technologies to improve productivity in multiarchitecture data centers and private clouds. It also includes the zEnterprise BladeCenter Extension (zBX), which allows for tight integration with IBM Smart Analytics Optimizer, IBM WebSphere® DataPower® and IBM Power® processing blade elements. Further, through use of IBM Tivoli® management and provisioning products, IBM provides the ability to manage an existing set of architectures that include standalone computing as well as virtualized computing resources.

This paper shows how today's organizations can leverage System z and its database technologies in the cloud, describing the benefits that System z users can gain by utilizing private cloud capabilities within their enterprises.

Defining "cloud"

Cloud has a long—and growing—list of definitions. The National Institute of Standards and Technology (NIST) defines it as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources…that can be rapidly provisioned and released with minimal management effort or service provider interaction."[1] The training organization Global Knowledge says cloud "can be broadly defined as several different methods to deliver information or services," and also describes it as "a new way of doing business that takes advantage of building efficiencies into the system that can scale out to provide services for multiple companies."[2]

As illustrated in Figure 1, there are various cloud models that organizations can build; the choice depends on the organization's business model. At the most basic level, infrastructure as a services (IaaS) provides computing resources as a virtual set of resources that are shared and dynamically provisioned. Platform as a service (PaaS) builds on this model, providing application middleware (including databases and development tooling) as blank canvases ready for the deployment of a specific application. Database as a service (DBaaS), the subject of a considerable part of this paper, can be considered as a specialized PaaS. Software as a service (SaaS) and business process as a service (BPaaS) enable higher levels of service delivery based on specific customer-written applications or business processes.

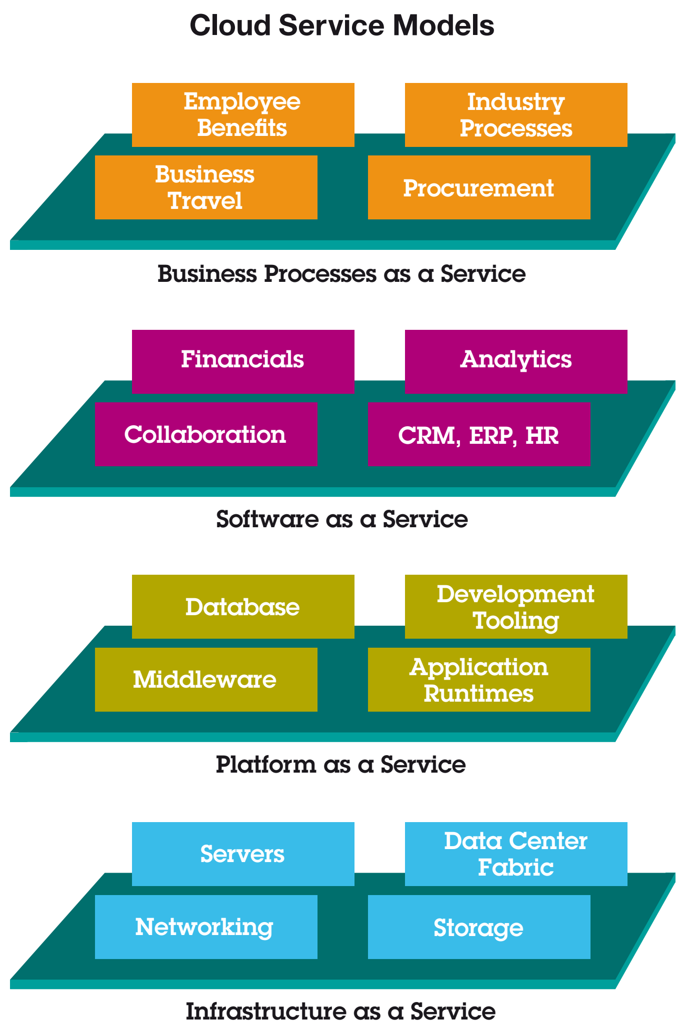

Figure 1: Cloud services that can contribute to a comprehensive services portfolio

As shown in Figure 2, the services provided in a cloud computing environment go through six different stages:

1. Creating the service offering
2. Offering the service to customers
3. Service subscription
4. Deploying the service
5. Managing the service
6. Terminating the service when it is no longer needed

Looking forward, the often complex tasks of building a service can become a self-service automated process in the progression to the full capability of what cloud can offer.

Figure 2: Cloud service life cycle management

It's also important to remember that services need to be made available within the required security framework, with accountability for costs such as CPU and software license consumption. Cloud-based solutions can transparently leverage the qualities of service that System z offers, making System z a highly available, secure, resilient and scalable platform for cloud computing.

Leveraging database assets in the cloud

Over the past several years, IBM has been developing and acquiring a set of products that enable organizations to further leverage their database assets in a manner that allows for increased business value from those assets. These products include:

- IBM SPSS for Linux on System z – Predictive Analytics
- IBM Cognos® Business Intelligence for Linux on System z
- IBM InfoSphere™ data tools
- IBM InfoSphere Optim® archiving tools
- IBM FileNet® document handling
- IBM Enterprise Content Management
- IBM ILOG® application integration tools

The depth and breadth of technology offered by these products, when integrated with application tools such as IBM WebSphere Portal and WebSphere Process Server, enable frameworks that allow organizations to operate more efficiently. Creating and implementing these complex relationships and applications can be a daunting task, and that is where introducing cloud into the picture can help. Operating in the cloud not only reduces complexity, but also brings these capabilities into real-time use to help increase operational efficiency. Consider the following examples from the insurance industry.

Example 1

An insurance company wants to improve their claims adjudication process. To do this, they have decided to use WebSphere Process Server to build the actual flow of how the claims are to be processed. As the system grows to improve the adjudication process, a growing number of WebSphere process servers are required by developers and, ultimately, the claims processing staff. Building process servers by hand to meet this need can take days; with automation through the cloud, this can be reduced to less than an hour. In addition, offering the creation of the process server as a service in the cloud enables a self-service approach to tasks that were previously performed exclusively by the IT staff.

Example 2

The insurance company's CIS system is deployed on an IMS database on z/OS. New SQL applications are deployed on a WebSphere Application Server instance that is in the cloud. The company decides to make WebSphere Application Server available through the cloud in order to facilitate rapid provisioning across a set of consumers in a standardized process. Furthermore, the company wants SQL applications to have direct access to the CIS data. Using the IMS Open Database capability, WebSphere applications running in the cloud will have direct SQL access to the CIS IMS database on z/OS.

Example 3

Building on the claims process above, the company is deploying SPSS to begin deeper analysis on an ongoing basis. Financial analysts within the customer site are being served SPSS directly to their desktop applications, and the SPSS servers are being delivered by the cloud infrastructure. Without the cloud, the SPSS servers would have to be provisioned by the IT staff on behalf of the financial analysts.

Example 4

Further data analysis work is being done through development and deployment of fraud analytic tools. These tools can also be made available via a cloud, with the goal of setting up notification processes to quickly alert management to potential fraud and thereby reduce the fiscal and business impact of fraud. Adding these tools to the cloud infrastructure provides flexibility and self-service to both the developers and the fraud detection department.

Aspects of the preceding examples can be found in most organizations today, where requirements for application integration and new development systems can strain IT resources. Without a cloud platform to work from, islands of computing and analysis can add greatly to complexity and costs—especially ongoing costs like the overhead of moving data from one source to another so that yet another system can begin using it.

Database as a service (DBaaS)

One of the most compelling cloud service models in Figure 1 is PaaS, particularly when IBM database technologies are on offer. These are useful offerings for cloud-deployed applications that require database resources and therefore need to have a database provisioned on demand. Ordinarily, database provisioning can be an onerous process that requires completing requisition forms and involves multiple teams from areas such as application development, system programming and database administration. The goal of database as a service (DBaaS) is to provide automation for database provisioning so that all required authorization, machine resources, database objects and connection information can be obtained seamlessly. As shown in Figure 3, the application selects the required database features from the cloud and gets database-connection information in return.

Figure 3: Providing automation for database provisioning

DBaaS topology

Cloud support for a database is multilayered. The layers include the application, the cloud manager and the database. The application layer views the other layers as abstractions and merely requests the services of a database. The cloud manager layer is tasked with finding an appropriate database management system (DBMS) to use and creating the needed objects. Ultimately, it will return information back to the application layer so it can begin using the database.

The cloud manager layer controls a repository of information regarding available DBMSes. It gathers information from the application layers (such as what type of service is being requested, and what is the expected load). It can apply the dynamic information provided by the application layer to static profiles to speed up database definition. It then searches available DBMSes, finds an appropriate match, defines the database and returns connection information to the application layer.

The database layer must support the basic functions of a cloud manager and provide functions to help isolate applications within the cloud. The most critical requirements for a cloud-enabled database are as follows:

- Provide isolation between cloud-enabled applications. There are four key levels of isolation: Data isolation ensures that each application in the cloud is not authorized to access other applications' data. Namespace isolation ensures that applications using the same database objects and runtime object names do not collide. Load/performance isolation ensures that applications can achieve performance goals. And failure isolation insulates each member of the cloud from failures of other members.
- Provide real-time load/use information so an appropriate DBMS can be selected by the cloud manager, and provide real-time load balancing.
- Scale to support a large number of cloud users.
- Reduce costs for each application deployed in the cloud.

DB2 for z/OS: Supporting the requirements of DBaaS

DB2 for z/OS has the availability and scalability features that make it the database of choice for many organizations' business-critical applications. As a highly scalable database that can run varied applications at high utilization levels, it is a natural choice for cloud deployment. DB2 for z/OS in a cloud provides a unique opportunity to take advantage of System z availability and scalability for all applications. DB2 for z/OS specifically offers robust support for DBaaS and satisfies the critical requirements of DBaaS in the following ways.

Meeting DBaaS data isolation requirements

- Data-level isolation can be achieved by standard DB2 access controls. DB2 natively supports authorization control and can also use authorization facilities like RACF®. DB2 10 for z/OS also supports powerful row and column access control, making it a great choice for applications requiring different data views for end users.
- Namespace isolation is available for both runtime SQL packages and database objects. Runtime SQL packages can be isolated via package collections. For example, application A could have packages pkg1 and pkg2 bound into collection A, while application B could have the same packages bound into collection B. The collection allows for application B to make changes to the underlying packages without affecting application A. Database objects can be isolated by object qualifiers. Application A can create a table APPA.TABLE1 and Application B can create a table APPB.TABLE2. The cloud manager can help guarantee the uniqueness of collection identifiers and object qualifiers.

- Load and performance isolation can be achieved by system configuration or database isolation. Hardware resources such as CPU and real memory can be managed robustly. DB2 for z/OS uses the z/OS Workload Manager to control access to CPU resources on a thread basis. Thus, multiple applications and users can be sharing a given DB2 for z/OS subsystem and meet differing CPU utilization requirements. Real memory can be controlled at the subsystem level in DB2 10 for z/OS. Many DB2 subsystems can be deployed on a single z/OS image and real storage encroachment can be prevented via DB2 configuration parameters. If strict performance isolation is required, a DB2 for z/OS instance can be deployed on a dedicated z/OS logical partition (LPAR). z/OS supports multiple LPARs per System z mainframe, so consolidation can still be achieved in such a configuration.
- DB2 for z/OS has robust fault tolerance and recovery. An application should not be able to cause a database failure, and all application errors (such as SQL errors, incorrect parameters passed, etc.) should be handled transparently to other applications.

Providing real-time load/use information

Real-time load/use information can be obtained directly from a DB2 for z/OS subsystem or group. DB2 provides CPU and storage utilization statistics via the Instrumentation Facility Interface (IFI) read synchronous (READS) interface. This data can be used in real time by the cloud manager to make an appropriate deployment decision. In addition, System z hardware provides opportunities to expand processing power in real time. LPAR configurations can be changed to reflect new memory allocation and processing power requirements. Additional processors and disks can be added as hot-swaps to the system.

Scaling to support a large number of users

DB2 for z/OS (in particular DB2 10 for z/OS) is a highly scalable system. Each DB2 subsystem can support a large number of concurrent users (5,000 to 20,000 in DB2 10 for z/OS). With varied applications deployed on a single subsystem, DB2 is able to run at a high utilization level while maintaining performance and stability.

Providing opportunities for cost reduction

The scalability of DB2 for z/OS provides a great opportunity to reduce costs through consolidation via the cloud. On top of the value created by hardware consolidation, additional cost benefits can be achieved from reduced configuration requirements. Some system resources such as buffer pools, system parameters and dynamic statement caches are shared for a given subsystem. This brings value to a cloud configuration as the management of these resources need not be done on a per-application level.

The amount of cost reduction achieved by DB2 for z/OS in a cloud-enabled application environment is largely dictated by the requirements of the application requesting the database services. For most applications, merging on a single DB2 for z/OS subsystem or data-sharing group can help cut hardware, maintenance and configuration costs. In addition, consolidation helps reduce the number of administrators needed.

IMS: Lowering risks and costs in the cloud

IMS is the high-performance application and data server for System z that offers a powerful combination of extremely high performance, scalability, rock-solid reliability and runtime efficiency. When it comes to running core applications that are at the heart of business processing, large corporations worldwide continue to depend on IMS. In the cloud, IMS leverages System z's support for cloud computing, as illustrated in Figure 4.

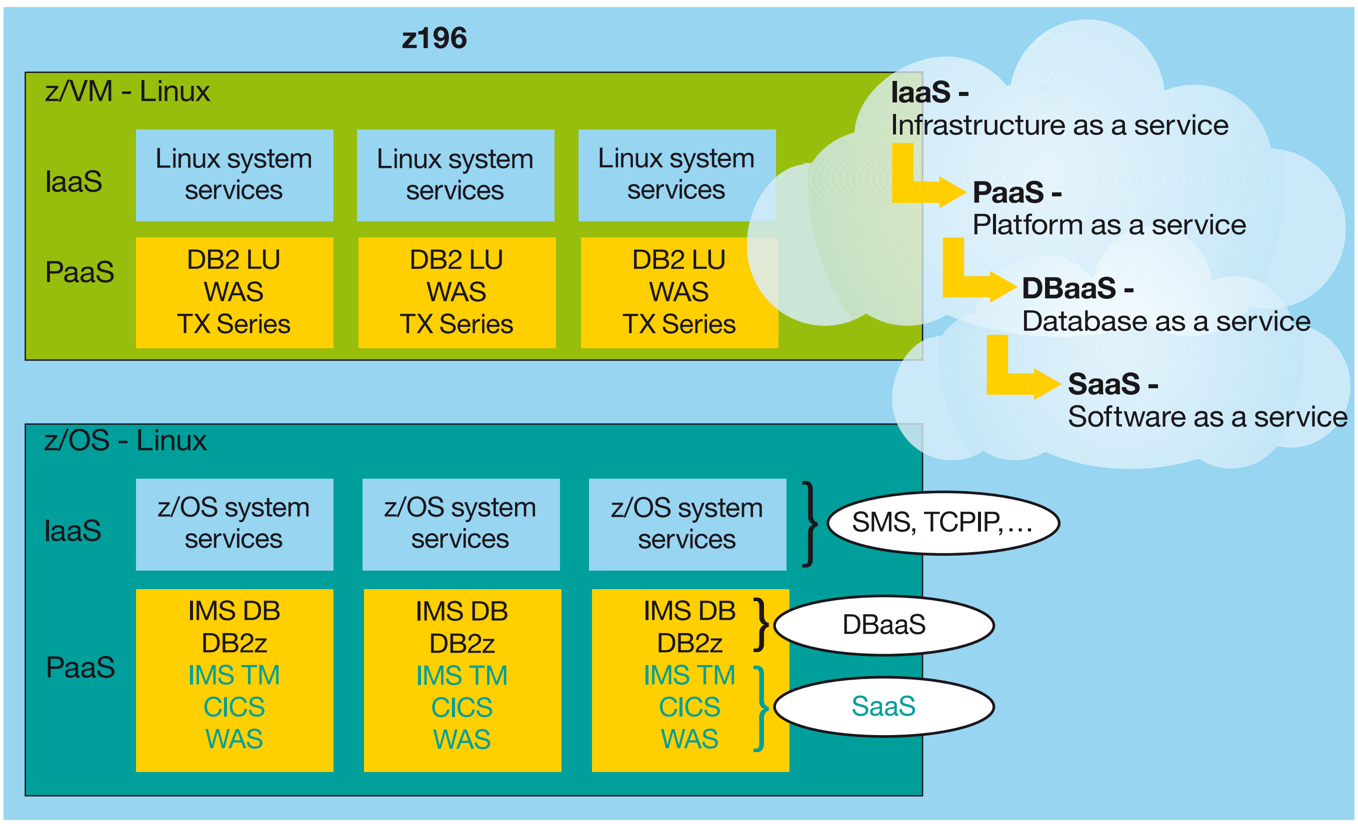

Figure 4: IMS dependence on System z support for cloud computing

IMS DB: Supporting the requirements of DBaaS

The database layer of IMS can provide the most critical requirements for a cloud-enabled database through the basic functions of the cloud manager.

Meeting DBaaS data isolation requirements

- Data isolation can be achieved by standard IMS access controls through various IMS database types such as Full Function, Fast Path, and HALDB. IMS natively supports authorization control and can also use authorization facilities like RACF. IMS supports both data isolation and database isolation by segment, area or partition.
- IMS can control the transaction workload within the PaaS architecture. The IMS Transaction Manager provides the means for dynamic control over DBaaS for application-to-segment sensitivity and application-to-record-data isolation within an area or partition.
- Load and performance isolation through IMS DB services uses clustering, partitioning and database isolation. In addition to what z/OS provides in virtualization, IMS uses its Base Primitive Environment subsystem to request system services and manage its participation in the use of real and virtual memory. Many IMS subsystems can be deployed on a single z/OS image. An IMS instance can be deployed on a dedicated z/OS LPAR. z/OS supports multiple LPARs per System z mainframe, so consolidation can still be achieved in such a configuration.
- IMS has robust fault tolerance and recovery. An application should not be able to cause a database failure, and all application errors should be handled transparently to other applications.

Providing load/use information

Load/use information is collected by IMS and z/OS through IMS logs and SMF/RMF data. This data can be used by the cloud manager to make an appropriate deployment decision. In addition, System z hardware provides opportunities to expand processing power in real time. Through the Hardware Management Console, LPAR configurations can be changed to reflect new memory allocation and processing power requirements. Additional processors and disks can be added as hot-swaps to the system.

IMS as a PaaS-complete solution

IMS is a highly scalable system. Each IMS system is elastic and can support a very large number of concurrent users. Application deployment can be scaled from a single IMS image to a clustered parallel IMSplex. IMS can run at a high utilization level while maintaining performance and stability.

As alluded to earlier, IMS is both a transaction manager and a database manager. The two work together seamlessly to provide a PaaS architecture for application containers, queues and data handling. The IMS platform controls workloads of various types:

- Transactional
- Data integration
- Loosely coupled
- Operational BI

IMS in total provides the control of the workloads and data so a cloud manager simply needs to support the operational services and business services of IMS to provide a robust PaaS solution with its associated DBaaS subset. Transactional workloads leverage IMS Database Manager and IMS Transaction Manager capabilities for multiple programming, scheduling, load balancing and authentication services. Loosely coupled workloads can be produced using IMS Connect, IMS Enterprise Suite SOAP Gateway, IMS Transaction Manager Resource Adapter, Web 2.0, and IMS Open Database.

PaaS tooling

PaaS tooling provides an integrated development environment (IDE) in which to create IMS transactions and databases as services in the cloud. This enables simplified application development for IMS using IMS Explorer for Development, which is currently available as a technology preview from **ibm.com**/ims. In addition, IMS Explorer for Administration can be deployed as an operational console with which to configure and manage IMS assets in the cloud.

Providing opportunities for cost reduction

The scalability of IMS provides a great opportunity to reduce costs through consolidation via the cloud. On top of the value created by hardware consolidation, additional cost benefits can be achieved from reduced configuration requirements. Some system resources such as buffer pools, system parameters and queues are shared. This brings value to a cloud configuration as the management of these resources need not be done on a per-application basis. IMS in the cloud enables organizations to develop and test complex system configurations with low risk and at low cost.

System z: Low-risk entry to cloud computing

With its strong security, high availability, virtualization capabilities and other features that can be critical to success in the cloud, System z offers an outstanding platform on which to build a private cloud. It also provides a solid basis for cloud deployment of database products such as DB2 and IMS and the business-critical applications that rely on them.

Summary

Cloud computing is a model of consuming and delivering IT and business services that enables users to get what they need, as they need it—from high-end, reliable database technology and business applications to IT infrastructure and platform services. Capabilities of self-provisioning and policy-driven management provide for faster deployment and faster time to production, and they help reduce IT staff involvement. Deploying DB2 on z/OS and IMS in a cloud environment provides significant economies of scale, while allowing a broader audience greater flexibility in exploiting key functions and the wide capabilities of the System z information management portfolio.

For more information

To learn more about how the strengths of System z can help your organization operate successfully in the cloud, visit: **ibm.com**/systems/z

Authors/contributors: Kershaw Metha, Peter Miotke, Mark Simmonds, Sandy Sherril, Bill Reeder, John Tobler

[1] Mell and Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, Information Technology Laboratory. http://csrc.nist.gov/groups/SNS/cloud-computing

[2] Baca, Steve, "Cloud Computing: What It Is and What It Can Do for You," Global Knowledge. www.globalknowledge.ae/knowledge%20centre/white%20papers/virtualisation%20white%20papers/cloud%20computing%20what%20it%20is%20an.aspx

IBM Systems and Technology Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
June 2011