Global Journal of Computer Science and Technology
Cloud & Distributed
Volume 12 Issue 11 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Data as a Service (Daas) in Cloud Computing [Data-As-A-Service in the Age of Data]

By S.Rajesh, S.Swapna & P.Shylender Reddy
University College of Engineering, Osmania University, Hyderabad

Abstract - Data has become the enabling technology for many of the recent innovations. "More data trumps smarter algorithms" has been the mantra behind this revolution in computing. Given the rate at which the data is produced, there is need for scalable solutions to extract information out of them. Allowing the data to be stored in the cloud and be accessed without geographical and scalability limitations will remove many bottlenecks in bringing data-oriented innovations. Current cloud architecture solves the issues of accessibility and scalability, but poses several new challenges such as automatic management of the service, pricing the data, and security of the data. This talk will include several techniques to address these challenges using automatic physical design, service-based pricing, and cryptographic mechanisms. Data ➔ Information ➔ Knowledge ➔ Intelligence.

Keywords : *Daas, Aas, Mashups Eai, Cio's, Gs1, Crm, Silos And Loe.*

GJCST-B Classification: *C.2.4*

Strictly as per the compliance and regulations of:

Data as a Service (Daas) in Cloud Computing [Data-As-A-Service in the Age of Data]

S.Rajesh [α], S.Swapna [σ] & P.Shylender Reddy [ρ]

Abstract - Data has become the enabling technology for many of the recent innovations. "More data trumps smarter algorithms" has been the mantra behind this revolution in computing. Given the rate at which the data is produced, there is need for scalable solutions to extract information out of them. Allowing the data to be stored in the cloud and be accessed without geographical and scalability limitations will remove many bottlenecks in bringing data-oriented innovations. Current cloud architecture solves the issues of accessibility and scalability, but poses several new challenges such as automatic management of the service, pricing the data, and security of the data. This talk will include several techniques to address these challenges using automatic physical design, service-based pricing, and cryptographic mechanisms. Data ➔ Information ➔ Knowledge ➔ Intelligence.

Keywords : Daas, Aas, Mashups Eai, Cio's, Gs1, Crm, Silos And Loe.

I. Introduction

Data as a service, or DaaS, is a cousin of software as a service. Like all members of the "as a Service" (aaS) family, DaaS is based on the concept that the product, data in this case, can be provided on demand to the user regardless of geographic or organizational Separation of provider and consumer. Additionally, the emergence of service-oriented architecture (SOA) has rendered the actual platform on which the data resides also irrelevant. This development has enabled the recent emergence of the relatively new concept of DaaS.

Data provided as a service was at first primarily used in web mashups, but now is being increasingly employed both commercially and, less commonly, within organizations .Traditionally, most enterprises have used data stored in a self-contained repository, for which software was specifically developed to access and present the data in a human-readable form. One result of this paradigm is the bundling of both the data and the software needed to interpret it into a single package, sold as a consumer product. As the number of bundled software/data packages proliferated and required interaction among one another, another layer of interface was required. These interfaces, collectively known as enterprise application integration (EAI), often tended to encourage vendor lock-in, as it is generally easy to integrate applications that are built upon the same foundation technology.

The result of the combined software/data consumer package and required EAI middleware has been an increased amount of software for organizations to manage and maintain, simply for the use of particular data. In addition to routine maintenance costs, a cascading amount of software updates are required as the format of the data changes. The existence of this situation contributes to the attractiveness of DaaS to data consumers because it allows for the separation of data cost and usage from that of a specific software or platform. Store data on cloud and provide results on the data as service. We can generate an astounding amount data. More data → smarter algorithm. Solves intractable problems like Automatic driving, Machine Translation, Semantic search like **Stone → Bronze →Iron →Oil →Computer → Data**

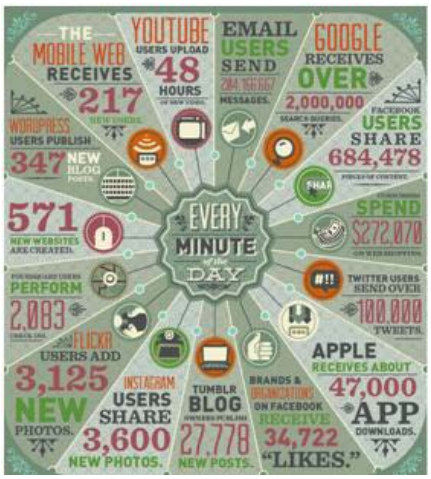

Figure 1 : Data disperse

As companies begin to decouple data from applications to enable richer services both internal and external to their organizations, new challenges arise that can both speed up as well as slow down adoption of data sharing using data-as-a-service offerings.

Recently Accenture formulate six predictions for game-changing technology trends. On the subject of the Industrial Data Services trend, it says that the **"freedom to share data will make data more valuable – but only if it's managed differently."**

Author α : Asst. Professor, Dept of CSE. M.Tech in Computer Science from University College of Engineering, osmania university, Hyderabad.
Author σ : Asst. Professor, Dept of CSE. M.tech in Vaagdevi Engineering College Bollikunta, Warangal.
Author ρ : Dept of CSE. Had 3 years of experience in teaching and 2 years in Industry. MCA from Osmania University, Hyderabad Andhra-pradesh India.

Figure 2 : Data sharing

Here these are eight dimensions of the value of data to an organization, which are difficult to measure in terms of actual prevalence, as follows:

1. Utility
2. Uniqueness or exclusivity
3. Ease of production
4. Usage and sharing restrictions
5. Usability and integration
6. Trustworthiness
7. Support
8. Consumer demand

The difficulty in shifting to an architecture that enables data as a service (DaaS) lies in a change in philosophy that CIOs have held for years. Who owns data? Is it the application, the application group, the organization, or …? To answer these questions, first consider the value of the data based on eight dimensions. However, to truly understand the value of DaaS and the shifting philosophy of data ownership to data stewardship, defining the data value chain is the first step.

II. Data Value Chain

The real value of data as a service can be measured by the length of the data value chain. When a company decouples data and makes it available for consumption as a service, they add value to the data value chain. The service can be used by the next "hop" to add value. The challenge is when consumers use the service to access the federated data and create their own isolated data pool silos. Currently, most data as services tend to look like a one-directional hub-and-spoke model.

Figure 3 : Data consumers request's/response model results in no data sharing

In this example, the data-as-a-service model is not being truly leveraged as a data value chain. Think of the example that Accenture gives where a company is grabbing data from its customer relationship management (CRM) system to study loyalty trends and, in the process, the marketing group creates its own isolated data pool, which is not usable by anyone else. What's even worse is if that data pool isn't constantly refreshed by the originating source. This model (Fig. 3) creates a break in the data value chain that causes data to quickly become old and inaccurate.

Alternatively, a healthy data value chain will have many consumers who may not continue the chain but won't store a snapshot view of the data they use. Instead, data will freely flow in and out of the core data sets as needed. The goal is to lay out the services in a way that produces a healthy ecosystem of data services along the value chain.

Figure 4 : Basic data value chain

Figure 4 shows the value of interweaving DaaS providers to increase the value of data and promote data reuse. The synergistic value comes when two or more services can feed each other value. This enables the ecosystem to grow the data value chain by way of enrichment between data consumers and data producers

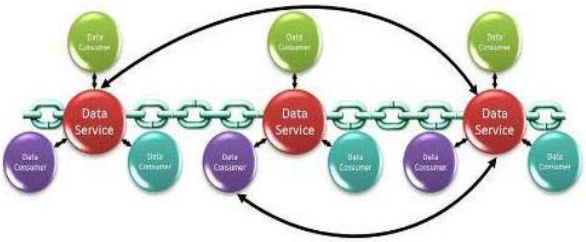

Figure 5 : Data value chain-added data enrichment

The challenge facing organizations is twofold:

1. To show the value in decoupling data from applications and
2. To create an open service-oriented architecture that allows the data to be shared both inside and outside of the organization.

III. Use Case Examples

Let's take a look at a couple of real examples of how a federated data-as-a service model can produce benefits not possible by traditional data management approaches.

a) Use Case 1: Secondary use of personal health data

An example of how setting data free can produce tangible benefits is illustrated by the sharing of health data via data-sharing services. The ability to decouple patient record data and aggregate it benefits medical facilities and practitioners by allowing them to:

- Monitor the health of the population
- Identify populations at high risk of disease
- Determine the effectiveness of treatment
- Quantify prognosis and survival
- Assess the usefulness of preventive strategies, diagnostic tests and screening programs
- Inform health policy through studies on cost-effectiveness
- Support administrative functions
- Monitor the adequacy of care.

Without decoupling and sharing the data via data services, these health benefits are extremely difficult, if not impossible, to achieve.

b) Use Case 2: GS1 Global Data Synchronization Network (GDSN) benefits to suppliers

GS1 allows suppliers, customers, manufacturers and producers to publish and synchronize product data into global data pools. Accenture and Capgemini have published reports describing the business benefits of GDSN based on extensive research with major associations, suppliers and retailers including Royal Ahold, The Coca-Cola Company, General Mills, The Hershey Company, The J.M. Smucker Company, Johnson & Johnson, Nestlé, PepsiCo, Procter & Gamble, Sara Lee, The Gillette Company, Unilever and Wegmans.

The results clearly show that synchronizing accurate and properly classified data brings many business benefits to suppliers. For example, time-to-shelf was reduced by an average of two to six weeks, order and item administration improved by 67 percent, and item data issues created during the sales process were reduced by an average of 25 to 55 percent (source: http://www.gs1.org/gdsn/ds/suppliers).

The business benefits include:

- Better category and promotion management
- Easier administrative data handling
- Smoother logistics
- Increased operational efficiencies
- Increased customer satisfaction
- Improved bottom line

IV. Challenges

a) Evaluate data silos – LOE and cost barriers

Not all applications can have data decoupled and shared. An organization has to survey its applications and data sources and identify those that can be decoupled and those that can't. The level of effort required to decouple data must also be weighed against the estimated value of doing it.

In many cases, the value of retiring the data store can be evaluated. For example, if a company loads in and statically stores tax tables, this can be painful to maintain and costly if the data gets stale. Even if the data can be decoupled and shared across multiple systems that can benefit from such access, the company may be better served by abandoning the data store altogether and subscribing to a data service that can provide tax table information in real time.

b) Privacy concerns

When an organization decides to share data to other applications and services outside of the current application or departmental walls — whether it is to other departments within the organization or to external organizations — privacy becomes a big concern.

Looking at our example related to the sharing of healthcare data for secondary use, the federal government passed the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules to ensure patient privacy of shared data.

c) Security concerns

When a company decides to enable data services, security is another area of general concern. Who can access the data, and how? Limiting access implies access control, which needs to be managed. If the data is going to be exchanged, especially between networks, will it be secure, and if so, how?

In light of recent PCI DSS Level 2 compliance breaches (credit card data privacy), the movement of data and the risk of unwarranted access can be difficult to prevent without a solid security plan in place to protect the data and control access to that data.

d) Falling short of true value

In addition to the cost of making data available as a service, companies also need to evaluate the value of doing it by answering these questions: Will any other service or application benefit from the availability of the data in question? Does the decoupling of the data allow for an upgrade and retirement path for a legacy application? There are plenty of cases where the value simply isn't there, so the utility of free data needs to be carefully weighed.

e) Data governance

Publishing and subscribing to data services require data governance to ensure the accuracy of the information being shared. We must have confidence that the data we receive and the data we submit is validated and harmonized. For example, we trust that a page on Wikipedia is accurate because there's a process in place to ensure the information is verified and corrected to be accurate. Like Wikipedia, the data value chain has to have data governance built into it to ensure the data available for use is current, complete and accurate.

Therefore, individuals and organizations that participate in the data value chain as non-terminating links have the responsibility to be data stewards. A data steward maintains data quality by ensuring that data:

- Has clear and unambiguous data element definition
- Does not conflict with other data elements in the metadata registry (remove duplicates, overlaps, etc.)
- Has clear enumerated value definitions
- Is still being used (remove unused data elements)
- Is being used consistently in various computer systems
- Has adequate documentation on appropriate usage and notes
- Documents the origin and sources of authority on each metadata element.

While these practices are hard to enforce, the goal is to make sure that no single point in the chain is the authority on correctness, but rather that each link plays a role in ensuring that the data is good.

V. Benefits

Data as a Service brings the notion that data quality can happen in a centralized place, cleansing and enriching data and offering it to different systems, applications or users, irrespective of where they were in the organization or on the network. As such, Data as a Service solution provide the following advantages:

a) Agility

Customers can move quickly due to the simplicity of the data access and the fact that they don't need extensive knowledge of the underlying data. If customers require a slightly different data structure or has location specific requirements, the implementation is easy because the changes are minimal.

b) Cost-effectiveness

Providers can build the base with the data experts and outsource the presentation layer, which makes for very cost effective user interfaces and makes change requests at the presentation layer much more feasible.

c) Data quality

Access to the data is controlled through the data services, which tends to improve data quality because there is a single point for updates. Once those services are tested thoroughly, they only need to be regression tested if they remain unchanged for the next deployment.

VI. Requirements for DaaS

- Cloud-enabled elastic,
- scalable, high-performance
- No need for CapitalExpence
- All costs must be Operating Expence
- No infrastructure related decisions needed

a) Data must remain secure

- Not only in-wire, but also at-rest

b) Automatic Configuration of DaaS

- Deciding on starting/shutting down nodes – elasticity.
- Deciding when to build/tear-down indexes – user experience.

c) Automatically Securing the DaaS

- Transparent to the user/provider
- In-line encryption/decryption
- Maintain full key control

d) Why automatic configuration?

- Reduce admin overhead
- Cloud converts Capital Expence into Operating Expence

e) E.Data Storage Services

i. *Advantages:*

- Data fragmentation and dispersal
- Automated replication
- Provision of data zones (e.g. by country)
- Encryption at rest and in transit
- Automated data retention

VII. Conclusion

The drawbacks of data as a service are generally similar to those associated with any type of cloud computing, such as the reliance of the customer on the service provider's ability to avoid server downtime. Specific to the DaaS model, a common criticism is that when compared to traditional data delivery, the consumer is really just "renting" the data, *using* it to produce a graph, chart or map, or possibly perform analysis, but for data as a service, generally the data is not available for download."Service Automation Units" (code that expresses the service interface) may contain methods for all "CRUD" operations (Create, Read, Update, Delete), as in traditional data operations, but data as a service is generally limited to Read.

Before a true revolution in data as a service can occur, organizations must be convinced of the value. While value of an IT change is traditionally measured in ROI, the benefits of decoupling data from applications for sharing across the extended enterprise are far-reaching benefits that can't always be quantified by sheer financial savings or gains. "Put simply, increased sharing of data through data services calls for a radical rethinking of how IT should handle data management. Essentially, data management shifts from being an IT capability buried within application support to a

collaborative effort that enables data to be used far beyond the applications that created it."

Is your organization a link or a break in the data value chain?

References Références Referencias

1. Machan, Dyan (August 19, 2009). "DaaS: The New Information Goldmine". Wall Street Journal. Retrieved 2010-06-09.
2. *Data as a Service: Are We in the Clouds?* 6. January 2010. 76–78. Retrieved 2010-06-09.
3. Dyche, Jill. "Data-as-a-service explained and defined". SearchDataManagement.com. Retrieved October 24, 2010.
4. "Statistical Data as a Service and Internet Mashups". Zoltan Nagy, United Nations. Retrieved 2010-06-09.
5. Cagle, Kurt. "Why Data as a Service Will Reshape EAI". DevX.com. Retrieved October 24, 2010.
6. "Data as a Service: Pricing Models for the Future of Data". Programmable web.com. Retrieved October 29, 2010.
7. Redkar, Tejaswi. "Chapter 4 – Windows Azure Storage Part I — Blobs". Windows Azure Platform. Apress, 2009.
8. http://searchdatamanagement.techtarget.com/answer/Data-as-a-service-explained-and-defined.
9. *http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-computing-v26.ppt)*.