2010 IEEE International Conference on Services Computing

Cross Enterprise Improvements Delivered Via A Cloud Platform: A Game Changer For The Consumer Product And Retail Industry

Trieu Chieu, Shubir Kapoor, Ajay Mohindra, Anees Shaikh
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598
e-mail:{tchieu,shubirk,ajaym,aashaikh}@us.ibm.com

***Abstract*—Gaining visibility into their retail supply chain has become a top priority for the Consumer Product (CP) industry. However, taking a "do-it-yourself" approach to the problem is proving to be both expensive and complex. Cloud Computing, with its on-demand provisioning capability on shared resources, has emerged as a new paradigm to address the challenges of the CP industry. In this paper, we describe a framework for deployment of business analytic solutions on a Cloud platform. We illustrate the benefits of the approach in context of the Demand Driven Business Analytic solution that provides demand signals to CP manufacturers.**

Keywords-Cloud; Demand Signal Repository;Supply Chain; Analytics; Composite Appliance.

I. INTRODUCTION

The phrase "Cloud Computing" [1-3] has become synonymous with containing and managing IT costs for enterprises. Cloud Computing is an IT delivery paradigm where compute capacity is made available to users in an on-demand fashion through a shared physical infrastructure. The expectation is that sharing hardware, software, network resources, and management personnel would reduce per unit compute cost for enterprises. Several vendors such as Amazon EC2, Google, and Rackspace have been offering commercial Cloud offerings. Though perhaps not enterprise grade level yet [13], Cloud Computing has gained the interest of several large enterprises, which have started deploying and experimenting with the technology for their test and development environments.

Current economic conditions and the need to lower costs have forced the Consumer Products (CP) industry to consider the cost advantages of Cloud Computing. According to a 2008 study of CIOs, the IT spend in most CP organizations is declining which demands a balance between controlling cost and providing business value. The industry typically loses 3.5% of sales ($40 Billion) annually due to supply chain inefficiencies. In distributed computing environments 85% of computing capacity is idle and 70% of the IT budget is spent on maintaining current infrastructure. Demand Driven Business Analytics (DDBA), one of the leading solutions that is gaining traction in the industry, focuses on a Demand Signal Repository (DSR) as a central information nucleus within the organization driving improvements across several business processes. The repository is characterized by the integration of demand data from retail partners, syndicated providers harmonized with internal CP supply chain information that drive analytics and optimization for intelligent decisions and insightful collaborations with trading partners. Due to the sheer volume of data, processing and integration expertise needed, an efficient implementation of a DDBA solution can easily exceed the budget for a CP organization. Additionally the workloads on the infrastructure driven by the data integration, analytic and end user routines are highly variable in nature leading to inefficiencies in infrastructure utilization. The emergence of Cloud Computing has opened up an opportunity to improve the cost structure, flexibility and efficiency in delivering a DDBA solution across multiple CP organizations.

In this paper, we present our experience in building a DDBA solution based on a Cloud computing platform, the benefits of which can be applied to both the CP organization as well as its retail customers. We first present a short overview of the DDBA solution along with challenges and requirements for a cloud-based solution. We then present the architecture and implementation of the solution, and discuss some issues that arise in applying a Cloud to DDBA.

II. OVERVIEW OF DDBA SOLUTION

Out-of-stocks on the shelf continue to run at 8%-10% for turn merchandise and 15-20% for promotions, resulting in loss of short-term sales and erosion of long term brand power and loyalty. Further, the uncertainty and variability caused by a lack of true demand information leads manufactures to stockpile extra inventory, thereby reducing the overall efficiency of a supply chain. There is a surge in the volumes of information available today and most organizations are drowning, struggling to understand what to do with the data they receive. The last mile blind spot i.e. visibility to retail point of sale data is rapidly being made available enabling consumer product manufacturers to manage their processes "outside-in" coordinated with their customer's demand requirements. The Demand Driven Business Analytics solution provides services managing information from shelf back to supplier. At the core of the solution is a DSR that is a robust, centralized database that stores, harmonizes and normalizes large volumes of demand data, including point-of-sale data, wholesale distribution information, inventory movement, promotional demographics, market demographics, third-party market

978-0-7695-4126-6/10 $26.00 © 2010 IEEE
DOI 10.1109/SCC.2010.35

content and customer loyalty data, to support better decisions in the areas of category management, joint value creation, VMI, trade promotion management, supply chain management and promotion management [11]. This is a leading edge solution quickly gaining traction within both the Retail and CP industries. With IT budgets razor thin and the perception for most organizations to focus on core business processes, creating and managing large volumes of information and investing in dedicated hardware is a thing of the past.

To address these problems we designed and implemented a cloud-based DDBA solution. During the design, we identified several requirements that had to be met by the design. These requirements are listed below

a) Reduce costs through multi-tenancy
b) Process and manage large volumes of data
c) Security of data and information
d) Rapid on-boarding of new retailers and CPs
e) Ability to plug different analytic providers
f) Custom reporting and ETL jobs.

A. Logical architecture of the DDBA solution

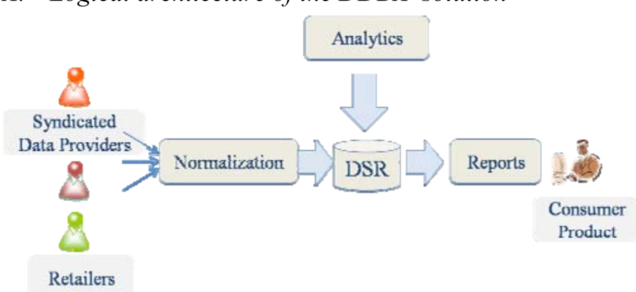

Figure 1 Logical architecture

Figure 1 shows a high level architecture of the DDBA solution. The solution aims to provide a collaborative platform for CP manufacturers and their associated Retail customers. The collaboration is initiated by Retailers and/or syndicated data providers securely uploading their daily Store / SKU level point-of-sale data to the solution. The uploaded data is cleansed, normalized and harmonized with internal CP master data attributes and persisted in the DSR. Based on the business objectives and challenges faced by the manufacturer, the solution allows for plugging in higher order analytics to improve and optimize business processes such as demand forecasting, new product introduction, on shelf availability, and order to cash [12]. The analytic results and corresponding key performance indicators and alerts are available via a role based enterprise dashboard for easy viewing and navigation. Figure 2 shows the sample output of a report from the solution.

Sort By: * Lagging Customer

As Of: 03-08-2010 **UOM:** CASE ● >=98.5% ● >=95% and <98.5% ● <95%

Customer	FY	Service Level	DOS	Total Sales	Total Outs	Avg On Hand	Avg On Order	Avg Sales
Mike Brothers Mart	2010	94.12%	10.51	2,810,091	175,550	108.13	37.39	10.29
HI Value Stores	2010	95.29%	7.53	6,378,335	315,214	125.50	109.48	16.66
McDermott Stores	2010	96.62%	19.65	6,666,308	233,127	179.36	52.68	9.13
Samuel Paul Corp	2010	96.69%	12.60	1,609,356	55,106	118.03	57.34	9.37
Henry Fine Quality Foods	2010	97.23%	15.23	11,110,009	316,584	93.21	35.96	6.12
WOW Foods	2010	97.70%	17.88	1,199,678	28,220	199.41	97.51	11.15
Hygiene Food Mart LLC	2010	97.73%	17.89	4,104,783	95,390	104.06	35.76	5.82
Fine Market Corp	2010	98.34%	11.20	5,231,981	88,431	391.19	277.53	34.94
Dollar Mart	2010	98.48%	9.64	11,796,118	182,210	107.57	53.18	11.16
Robert & Co Grocery	2010	98.57%	18.64	1,426,677	20,749	144.21	53.32	7.73

Top Page up Page down Bottom

Figure 2 Sample report output from the DDBA solution

The DDBA solution is a perfect candidate to be hosted on a Cloud platform as it exhibits several of the key workload attributes suitable for the Cloud bursty workload, compute intensive analytics, and the need to process large volumes of data.

III. CLOUD COMPUTING AND IMAGES

Cloud Computing promises lower costs due to better economies of scale. To achieve low cost, it is critical to eliminate manual processes associated with systems management and provisioning. In this section we briefly describe a high level architecture of a Cloud Computing model and discuss the role of images in the paradigm.

A. Cloud Computing

Cloud Computing provides a self-service environment for requesting compute resources. Figure 3 shows the logical architecture of a Cloud Computing deployment. Typically, service providers create a pool of networked hardware resources. Each hardware resource runs virtualization software such as VMWare, Xen, or KVM. Virtualization enables each hardware resource to host and run multiple virtual machines. The Cloud resources are made available to users through a User Portal or Web Services APIs. User requests are forwarded to a Provisioning Component that performs the following tasks:

1) Refers to a Resource Manager to locate a hardware resource that has available capacity to run the virtual machine that the user requested.

2) Copies the image for the virtual machine from the Image Library to the target hardware resource.

3) Creates the configuration for the virtual machine on the target hardware resource and creates the virtual machine.

4) Notifies the user after the virtual machine has been successfully created.

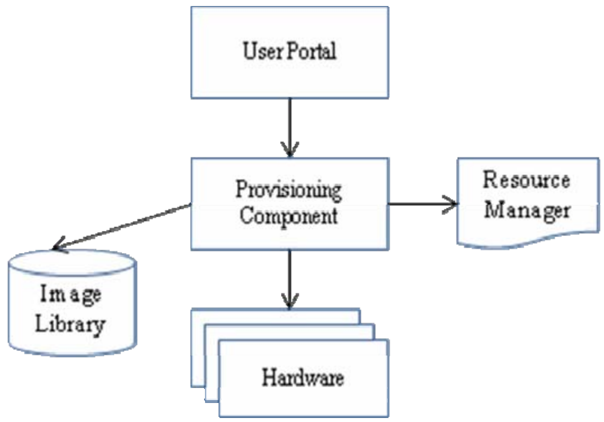

Figure 3 Logical Architecture of Cloud Computing

B. Images

An image is simply the disk representation of a virtual machine pre-installed with an operating system. A virtual machine image preloaded with application software is further referred to as an *appliance*. An image or an appliance consists of two files: the configuration file, and the actual disk image. The configuration file represents the metadata about location of the disk image file, display name, attached network and peripheral devices.

Cloud Computing uses images as the building blocks for provisioning. When a user requests a compute resource, the Provisioning Component locates and retrieves the appropriate image from the Image Library, and uses the image to create the new virtual machine. The capabilities provided by the Cloud are abstracted in the Infrastructure-as-a-Service layer.

IV. ARCHITECTURE OF A CLOUD-BASED SOLUTION

We architected the cloud-based solution to provide an affordable DDBA solution for CPs. The main areas of focus were multi-tenancy, service automation, and security. For version 1 of the solution, we chose to use platform-level multi-tenancy in which a dedicated instance of the solution was created for each customer, and the solution instance shared the underlying middleware among multiple CPs. To eliminate labor costs, we extensively used automation script to automate the customer on-boarding processes. In addition to traditional VLAN-based security, we designed the solution to securely isolate data, reports, ETL jobs, and DSRs for different tenants.

We discuss the architecture of the cloud-based DDBA solution in two parts. First, the Service Manager is responsible for the provisioning of the DDBA solution instance, adding/removing tenants and retailers, and performing change management operations on the IT infrastructure. Second, a set of Common Services, which are shared across multiple instances of the solution. These services include the Data Service that facilitates receipt of data from the retailers, and transferring it to the correct location for processing; the Authorization Service that manages userid information about different tenants; and the Analytics Service that selects a specific analytic component for use by the solution.

Figure 4 shows the architecture of the cloud-based solution. The Service Manager is used by a System Administrator. The three Common Shared Services are also underlined. The solution assets comprise of an Information Server, a Reporting Server, and a Database Server. All the servers are deployed on a Cloud provided by a standard Infrastructure-as-a-Service (IaaS) platform such as Amazon EC2 [5] or the IBM Developer Cloud [6]. IaaS provides REST APIs to create and delete virtual servers from images that were created specifically to support the solution. In this section, we discuss the three functional components of the architecture.

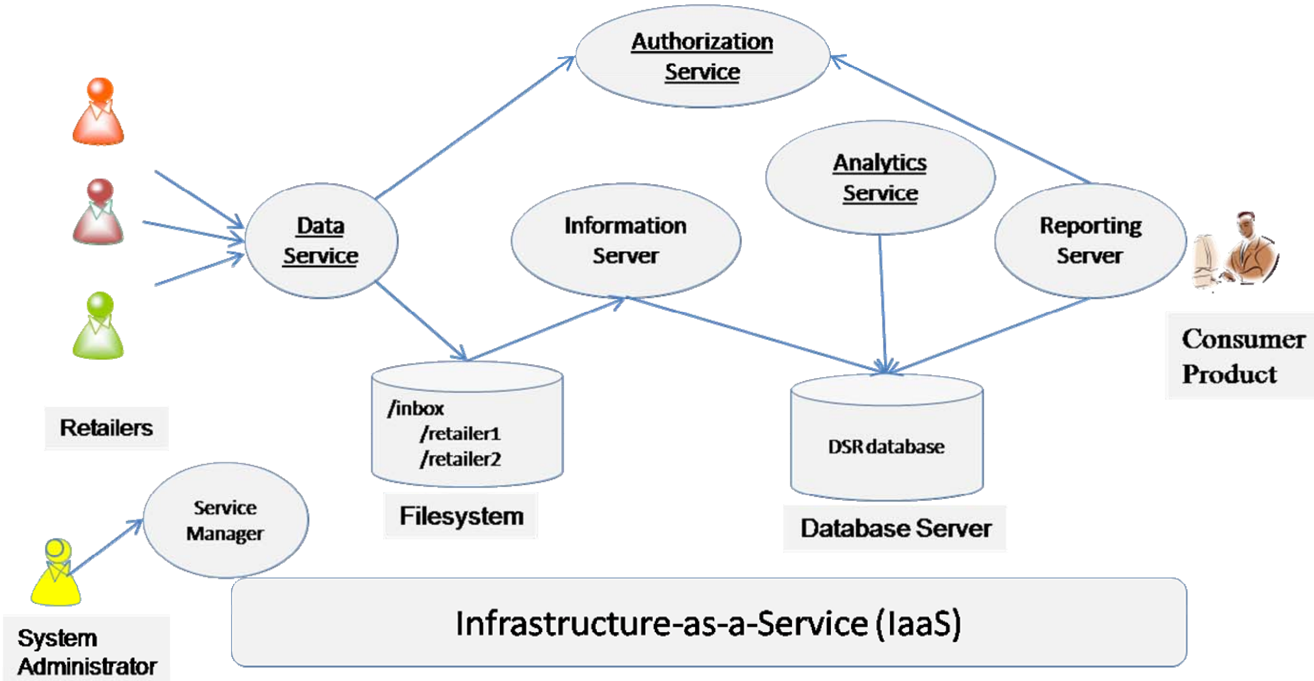

Figure 4 Architecture of the cloud-based solution

A. Service Manager

The Service Manager provides a service management abstraction layer on top of the infrastructure services in the Cloud. In our current implementation, the Service Manager has four main sub-components:

- a *service catalog* that provides a browseable view of solutions that are available for deployment
- a *service template* framework that allows administrators to create *service definitions*, that serve as business-level interfaces to present to Cloud users, and *service mappings* which specify how business level inputs are used to provision and configure services
- a *mapping engine* that converts user requests into workflows which set applicable parameters and invoke the appropriate provisioning APIs, Cloud services, or automations
- a *service portal* that renders the service definitions and allows users to select and deploy solutions from the service catalog

Figure 5 shows the components and logical architecture of the Service Manager. The service catalog database contains the structure and contents of the solution catalog which is displayed in the service portal. In addition, the database contains service templates for available solutions, each consisting of the corresponding service definition and service mapping. Both of the service template elements are created by the service provider, who is expected to know how the service should be represented to users, and also how inputs should be used in deploying the service. The Service Manager provides a framework for expressing these in a structured and standardized way.

1) Solution Deployment

The service template-definition consists of service mappings that point to the deployment topologies used to instantiate a solution.

The framework enables system administrators to specify a solution's requirements and deployment topology using an XML file (We are in the process of converting the XML to an OVF compatible format). The deployment topologies are represented as Composite Appliances. A Composite Appliance is a collection of individual appliance images that are preconfigured to work together. Though the detailed knowledge of configuration points in a Composite Appliance for a given solution is required, the requirement analysis and the determination of configuration points are beyond the scope of this paper.

A Composite Appliance is specified using an XML file. Figure 6 shows the UML diagram for specifying a solution requirements and topology of a Composite Appliance. The top-level class for the specification is an Appliance section. Each Appliance section consists of one or more Node sections. Further, each Node section consists of an Image Requirement section and an Image Specification section. The Node section corresponds to one logical node of the solution. The Image Requirement section specifies the memory, CPU, disk and network requirements of the image associated with the solution. This approach allows a system administrator to quickly configure and deploy different sized solution by merely changing the values in the specification.

The Image Specification section contains information about the image that is part of the solution. The information is used by the Provisioning Component at the time of deployment of the solution. As mentioned in Section III.B, each image may contain a preconfigured software stack. The values correspond to the following attributes

- OSType – Descriptive name of the Operating System name and version installed on the image
- ImageId – Identifier of the image in the image library
- Filename – File name for the image
- VirtualizationType - Type of virtualization technology needed for this image
- RunOnceScript – A semi-colon (;) separated string of commands that are executed when the newly-created virtual machine from the image is first started. We discuss this attribute in the following sections.
- Userid – Specifies the "Userid" with root privileges for the new virtual machine
- Password – Specifies the password that needs to be set for the "Userid" with root privileges

Each Image Specification section also contains multiple Software Attributes. The Software Attributes contain the name-value pairs describing the attributes associated with the software stack that is configured on the specific image. For example, in an image that has the DB2 Server software installed, the Software Attributes section would contain the db2port, db2InstallationPath, db2UserId, and db2Password information. Since an image may be authored by another person, this information provides system administrators with information about the image so that they can write scripts to change the configuration as needed. Additional details can be found in [4].

2) Service configuration and data onboarding

The Service Manager provides capabilities to request DSR-related services from the Cloud, including requesting a new instance of the DSR solution for a new CP tenant, adding and removing retailers associated with a specific CP instance, and performing change operations on an instance of the solution. As mentioned earlier, each of the available DSR services has a corresponding service template that allows the service creator to define how users interact with a service (service definition), and the set of actions to deploy, configure, and populate the service (service mapping).

When a user selects the "Add New Retail Data Source" service from the catalog, for example, the Service Manager decodes its service definition specification and renders it as a dialog in the service portal. The service definition enables service creators to easily create flexible dialogs for collecting user inputs at the desired level of abstraction. This is in contrast to existing infrastructure Cloud offerings in which the user is forced to describe his needs in terms of number of

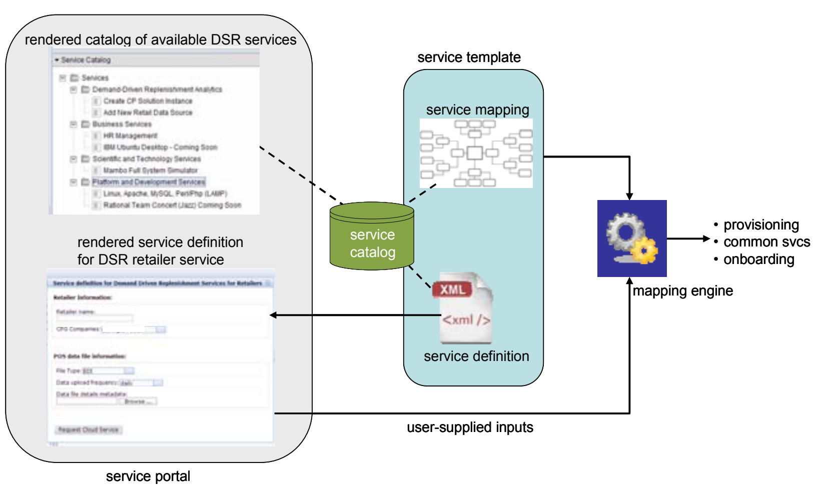

Figure 5 Components of the Service Manager

virtual CPUs, amount of memory or disk, software image, etc.

In this case, the service definition specifies information about the retailer (e.g., name, id), the CP tenant to which it wishes to make its data available, and information about the format of the data that will be uploaded (e.g., file type, the frequency of data upload, the number of product SKUs expected to be in each file, etc.). The service definition XML schema supports a number of ways to collect input from the Cloud user, and we have used it to create a variety of Cloud services, including analytics services such as the DSR, business-oriented services such as enterprise business process applications and virtual enterprise desktops, and infrastructure services such as configurable application development platforms.

The service mapping specifies how input parameters should be used in deploying the service. In general, the mapping may be as simple as a pass-through of an input parameter to a configuration script in a virtual machine instance, or a more complex transformation that considers a workload parameter (e.g., expected number of stores per retailer, or number of analytics report users), and sizes and configures a software component appropriately. For the retailer onboarding service, the information about the uploaded data is used to invoke the common Data Service in the Cloud (see Section IV.B.1) to establish an appropriately sized inbox for the retailer with the right set of authorizations.

The mapping engine combines the user-supplied inputs with the service mapping to execute the service deployment actions. When deploying a new solution instance on behalf of a CP tenant, the mapping engine requests the provisioning of a Composite Appliance as described in Section IV.A.1. For retailer onboarding, there is no virtual image provisioning required – instead, the mapping engine invokes the Data Service as described above, and also executes additional automation to set up ETL jobs in the Information Server to perform the required data transformation for the new retail data source. Similarly, if the CP tenant wishes to add new reports to its analytics (or remove a report), the Service Manager collects the report selection and any associated parameters, and then uses automation scripts to reconfigure the Reporting Server in the tenant's existing DSR solution.

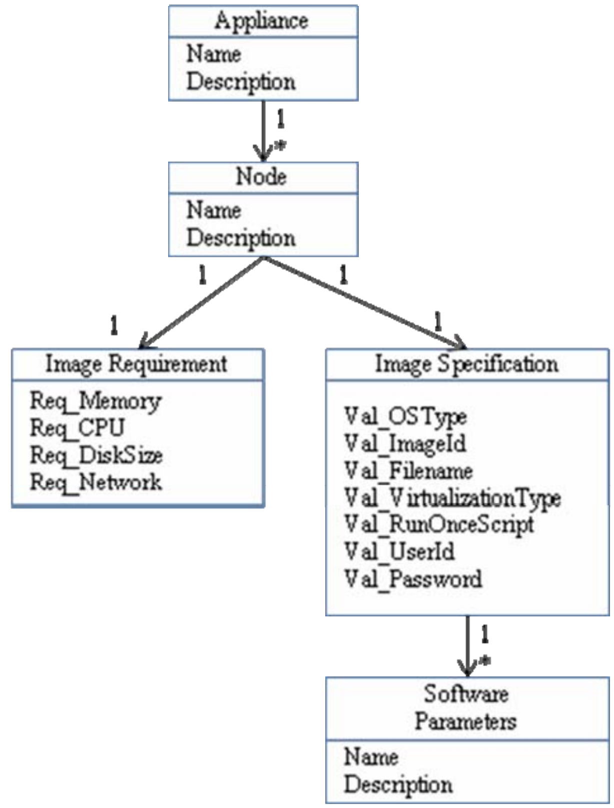

Figure 6 UML Schema of a Composite Appliance

B. Common Shared Services

1) Data Service

DDBA is a data-centric solution as it requires data to be uploaded to the solution on a periodic basis. As part of the solution, retailers need to upload their daily point-of-sale data. Based on the number of SKUs being processed, the datasets can range from 30 MB to 8 GB sized files. The role of the Data Service is to provide a bulk data transfer capability to receive data from each of the retailers and transfer the contents to the appropriate inbox. Even though bulk data transfers are best suited using FTP servers, we did not choose this option due to two reasons. First, it is difficult to validate whether the data file has been correctly uploaded. Second, there is no mechanism available as part of the FTP Server processing to trigger action upon successful transfer of the file.

The Data service provides a set of REST APIs that allow a system administrator to define users authorized to use the service and associate a valid inbox location (e.g., server name and folder, or an NFS mount point) to where the file should be transferred. The Data Service uses another shared service called the Authorization Service to authenticate users. When the Service Manager adds a new retailer to a solution as part of the retail on-boarding scenario, it configures the Data Service with a new user id and specifies the location of the inbox to where files uploaded by this retailer should be transferred to.

2) Authorization Service

The Authorization Service provides a common service for managing userid information about different customers in the Cloud. At the time of customer on-boarding, the Service Manager defines new userids and roles in the Authorization Service. The information and roles are also used by the solution instance to control access to the reports available to customers. The Data Service uses the Authorization Service to validate and control access for retailers to upload point-of-sale data.

3) Analytics Service

A design principle for the DDBA solution was to enable a CP manufacturer to leverage existing IT investments, i.e. vendor solutions or the ability to utilize pre-packaged analytics included in the DDBA solution. The choice of the analytics is largely driven by needs of the internal business process, prior experience with the vendor, or information availability. For a cloud-based solution, we developed a framework where a customer would specify their choice of analytics, choose from a set of pre-defined providers and provide integration parameters for interfacing the analytics with the DSR. The Service Manager notes the customer selection and configures the Data Service appropriately. As the retailer data is uploaded to the Cloud, the Data Service transfers the transformed data to the inbox of the selected analytic provider. After the analysis is complete, the results are moved back to the DSR and analysis made available via reports on the Reporting Server.

C. Solution Assets

The solution assets provide the core business analytics functions for the DDBA solution. The assets consist of an IBM Information Server, a Cognos Business Intelligence Reporting Server, and a DB2 Database Server on separate appliance images. These server images are configured and networked together to provide a set of integrated services. The Information Server is responsible in converting and aggregating the input data from different retailers into a staging database, and in executing ETL jobs to generate a DSR data warehouse in the Database Server. The Cognos Business Intelligence Reporting Server is used to perform OLAP on the DSR data warehouse and to generate reports for an enterprise mash-up dashboard to be accessed by CP tenants.

V. Exploiting the Cloud

Deploying any solution on a Cloud platform is in itself not very interesting because Cloud provides resources similar to physical machines albeit at a cheaper price. The key innovation lies in the solution exploiting the dynamic elasticity capabilities of the Cloud platform. In this section, we discuss how our framework enables the solution to exploit the Cloud capabilities.

A. One-size deployment topology does not fit all

Not all CP customers have same requirements. The main differences lie in the number of retailers that participate in demand signal generation process, number of product SKUs in each data file, and number of reports that a customer signs up for. As a result, we designed the solution to gather CP requirements at the time of on-boarding. The business requirements are mapped by the Service Manager to deployment topologies with specific configuration parameters. The configuration parameters reflect the CPU, memory, and disk sizes allocated to each of the virtual machines that are part of the solution. The underlying Cloud environment then creates the solution instances based on the specified configuration. The framework eliminates the need to over-allocate resources to the solution when the resources are not required.

B. Managing data

As we mentioned earlier, DDBA is a data-intensive solution. In steady state, the size of the DSR repository can grow up to 15 TB. Further, each participating retailer uploads point-of-sale data on a regular (hourly or daily) basis. The size of the data files can vary from 30MB to 8GB. These files are received by the Data Service component and moved to the inbox of the solution instance. To prevent out-of-space situations from occurring either on the inbox file-system or in the database server, the framework provides a mechanism where the Service Manager can set thresholds and subscribes to alerts from the appropriate virtual machines. These alerts are linked to specific corrective actions that the Service Manager initiates to remedy the situation. At present, we are only monitoring situations that could adversely impact the operations of the DDBA solution.

C. Managing growth

After signing up for the DDBA solution, we expect the needs of the solution to grow as enterprises increase their market reach and products that they manufacture. To accommodate growth, the Service Manager provides a change request mechanism. The mechanism allows a system administrator to add and remove retailers, and change cpu, memory resource allocation to active instances of the solution. The mapping engine of the Service Manager also automatically determines if on-boarding additional retailers would result in generation of additional workload that cannot be handled by merely adding more resources to the deployed instances. In this case, the Service Manager is designed to trigger horizontal scalability of the solution middleware (e.g. adding more application servers).

VI. Related Work

Current work in Cloud Computing has been centered on providing single VM Compute resources to the end users. Both, Amazon EC2 [5] and IBM Developer Cloud [6] provide the ability for users to provision single images in the

Cloud. Dynamic scalability of web applications deployed through appliances in a Cloud has also been addressed in [7]. No support is available to define and request multi-server solutions, or provide a set of common services for use by different solutions. Other vendors are focusing more on providing application runtimes (Google App Engine[8]), or delivery of specific business function via the software-as-a-service delivery model (Salesforce [9], GoogleDocs[10]).

VII. CONCLUSION

We have presented a framework for deploying business analytics solutions on Cloud platform. The solutions are characterized by a need to process and manage large volumes of data, rapid on-boarding of new retailers and CPs, and an ability to plug-in different analytic providers. The framework provides a standardized mechanism to deploy solutions in the Cloud. Our work has demonstrated the advantages of using a shared services based approach to manage costs and complexity of solution deployment for our customers. As part of future work, we plan to extend the work to enable system administrators to specify operational characteristics of the solution such as deployment zones, and availability requirements.

REFERENCES

[1] G. Gruman, "What cloud computing really means", InfoWorld, Jan. 2009.

[2] R. Buyya, Y. S. Chee, and V. Srikumar, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities", Department of Computer Science and Software Engineering, University of Melbourne, Australia, July 2008, pp. 9.

[3] D. Chappell, "A Short Introduction to Cloud Platforms", David Chappell & Associates, August 2008.

[4] Trieu Chieu, Alexei Karve, Ajay Mohindra, Alla Segal, "Simplifying solution deployment on a Cloud through composite appliances", In The Sixth International Workshop on System Management Techniques, Processes, and Services (SMTPS), April, 2010

[5] Amazon EC2, http://aws.amazon.com/ec2

[6] IBM Developer Cloud, http://www.ibm.com/cloud/developer

[7] T. C. Chieu, A. Mohindra, A. A. Karve and A. Segal, "Dynamic Scaling of Web Applications in a Virtualized Cloud Computing Environment", Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE 2009), Macau, China, Oct. 2009, pp. 281-286,

[8] Google App Engine, http://code.google.com/appengine/

[9] Salesforce, http://www.salesforce.com

[10] GoogleDocs, http://docs.google.com

[11] 2009 Sales and Marketing Report: CGT and AMR Research.

[12] Advances in Demand Signal Repositories for Consumer Products, AMR Research , Nov 2007.

[13] Kunwadee Sripanidkulchai, Sambit Sahu, Yaoping Ruan, Anees Shaikh, Chitra Dorai, "Are Clouds Ready for Large Distributed Applications?," *ACM SIGOPS Int'l Workshop on Large Scale Distributed Systems and Middleware (LADIS 2009)*, October 2009.