Volume 5, Issue 2, February 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Web-Services Testing Operational Disputes as Well as Concerns of Big-Data in the Cloud

Sanjay Ram M*, Alanthuraiyan V
Department of MCA & Adhiyamaan College of Engineering (Autonomous),
Hosur, Tamil Nadu, India

Abstract— Big-data is a key enabler of exploring business insights and economics of services. Cloud computing and big data are two fast-growing areas of IT. Big-data occupies a massive duty of hardware and processing resources, making approval costs of big-data technology irrational to small and mid-sized businesses. Cloud computing offers the guarantee of big data implementation to small and mid-sized businesses. Web Services are the foundations of modern distributed purposes. The ease and flexibility of integrating disparate components has made rapid application development possible across languages, platforms, devices and operating systems. Web-services testing (WST) especially when integrating to services owned by further retailers. They have to pact with the range of implementation techniques used by the further services and to gather an ample range of experimental requirements.

Keywords— End user big-data, Web-service testing, Operational dispute, numerous ways and Cloud Computing.

I. Introduction

Big-data pass on to the enormous amounts of data that collect over time that are difficult to evaluate and handle using general record management tools. Big-data includes commercial connections, e-mail messages, photos, scrutiny videos and activity logs. Scientific data from sensors can achieve infinite proportions over time, and Big-data in addition include formless text posted on the Web, such as blogs and public media [1].

Volume, variety and velocity are three defining assets or dimensions of big-data. Volume refers to the quantity of data, variety refers to the amount of types of data and velocity refers to the speed of data-processing. According to the 3Vs representation, the challenges of big-data management result from the growth of all three properties, rather than just the volume alone the absolute amount of data to be managed [2].

The big-data is range or the types of data being created. The 300+ billion emails per day can be considered structured text and one of the simplest forms of big-data. Economical transactions including movie ticket deals, gasoline deals, restaurant deals, etc., are generally structured and make up a small scrap of the data running around the global networks at present. Unstructured data is a primary source of growth in the variety. Music is an ever increasing variety of data and streaming nearly 19 million hours of music each day after the free music service, earlier [1][3].

The rising importance of big-data computing stanches from advances in many diverse technologies:

Cluster computer structures: The computer systems, consisting of thousands of "nodes," each having several processors plus disks, connected by high-speed local-area networks, has turned into the chosen hardware configuration for data-intensive computing systems [4]. These clusters offer both the storage capacity for huge data sets, and the computing power to organize the data, to reflect on it, and to respond to queries on the data since end-users. Evaluated with traditional high-performance computing (e.g., supercomputers), where the hub is on maximizing the raw computing influence of a system, cluster computers are designed to make the most of the reliability and efficiency with which they can handle and analyze very large data sets [5]. Cloud computing services: The rise of large data-centers and cluster computers has created a new business model, where businesses and individuals can charge storage and computing power, fairly than making the massive capital investments needed to rigid and stipulation extensive computer installations. For instance, Amazon Web Services (AWS) offers mutually network-accessible storage charged by the gigabyte-month along with computing cycles charged by the CPU-hour [6].

Data analysis algorithms: The infinite volumes of data require automated or semi-automated analysis methods to notice patterns, discover anomalies, along with extract information. Over again, the software algorithms fresh forms of computation, optimization, combining statistical analysis, along with artificial intelligence, are able to rigid statistical models from huge compilations of data additionally to gather how the system should react to new data. Scientific-data: Digital data are being generated by many different sources, including digital images (telescopes, video cameras, MRI machines), chemical and biological sensors (micro-arrays, environmental monitors), and even the millions of individuals and organizations making web pages. Computer networks: Data as of the several unusual sources can be collected into huge data sets through localized sensor networks, as well as the Internet. Data storage: Advances in magnetic disk technology have vibrantly decreased the charge of accumulating data. As a reference, it is predictable that if all of the text in all of the books in the Library of assembly could be converted to digital form, it would add up to barely about 20-30 terabytes. Service-Oriented Architecture (SOA) and its Web implementation Web Services (WS) promote an open standard-based and loosely coupled architecture for integrating applications in a distributed heterogeneous environment

[7][8]. Such applications are characterized by task distribution, service orientation, and collaboration between development parties, run-time actions and open standards for interfacing among their components [9]. A few years ago Web services were not fast enough to be interesting. Web services are application components; Web services communicate using open protocols; Web services are self-contained and self-describing; Web services can be discovered using UDDI; Web services can be used by other applications; HTTP and XML is the basis for Web services. Interoperability has Highest Priority: When all main platforms can access the Web using Web browsers, unlike platforms couldn't interact. For these platforms to work mutually, Web-applications be developed. Web-applications are basically applications to run on the web. These are built just about the Web browser standards and can be used by some browser on every platform.

II. OVERVIEW OF BIG DATA WITH CLOUD

A. Big-data on the Cloud

The term big-data is derived from the fact that the data sets are so bulky that typical database systems are not able to accumulate and analyze the data sets [10]. The data sets are bulky because the data is no longer conventional structured data, but data from many new sources, as well as e-mail, public media, along with Internet easy to get to sensors. The individuality of big-data presents data storage and data analysis challenges to businesses.

A distinctive model for in-house storage of big data is clustered Network-Attached Storage [11]. The configuration would start with a network-attached storage (NAS) pod consisting of numerous computers attached to a computer used as the (NAS) device. Several NAS pods would be attached to each other from side to side the computer used as the NAS device. Clustered NAS storage is an expensive prospect for a small to medium size business (Fig.1). A cloud service provider can furnish the necessary storage space for substantially lower expenses [12].

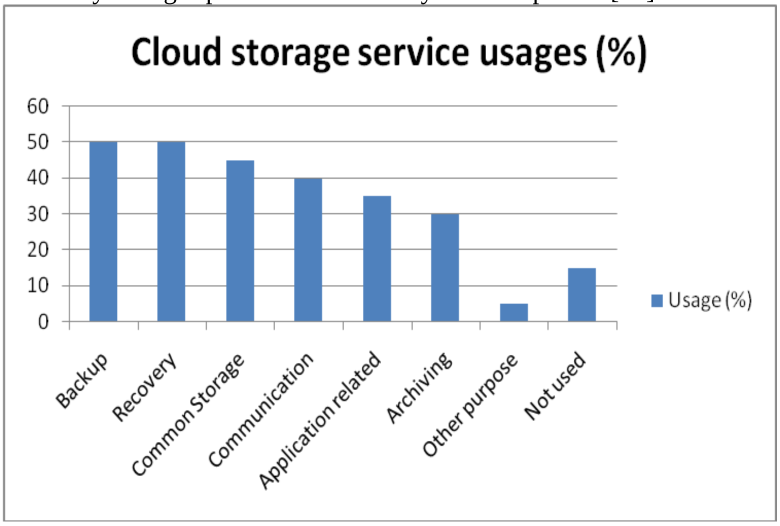

Fig 1: Cloud storage service usages for applications

B. Storage Systems to Cloud features useful for big data

The leading company is often engaged to help organizations address the challenges of peta byte scale harms. A new understanding of how storage and computing systems should be erected to achieve large scale. There are many drivers leading the industry away from traditional monolithic storage and computational approaches: *Cost,* traditional NAS, SAN, and RDBMS solutions are besides luxurious at peta byte scale. *Complexity*, there is no single conventional storage solution that can simply scale to meet the needs of these users. Multiple systems must be joined with external management components to reach the needed capability. *Management*, complex systems have greatly increased management overhead and are not easy to maintain. *Reliability*, the consistency requirement for multiple peta byte systems is unusual. Because no single storage device can hold all of the necessary data, various systems need to cooperate to offer reliability. Errors can frequently result in periods of data inaccessibility. In count, multiple peta byte systems are typically originated to span multiple data-centers. Conventional systems aren't erected to handle these matters.

C. Cloud architecture traits for storage systems

Cloud computing architectures, on the other hand, are characterized by their use of hefty quantities of reasonable, commodity systems with direct-attached storage, all working in concert to provide the user with a single system view that transcends the performance and ability of any single machine within the system.

A storage system built in this manner provides the following traits: Scalability, a cloud storage administrator is able to add extra computers, including their storage capacity, to a running system not including a loss of availability of files or administrative functionality. Capacity, the cloud storage system offers a global view or name space, aggregating the capacity of all attached storage devices. Reliability, cloud storage permits the user to specify how countless copies of each file to maintain in the system. The cloud is conscious of the loss of some machines in the system. When these errors take place, the cloud can alert the suitable administrators and take appropriate action to recover the requested reliability level. Geographic Distribution, an only instance of a cloud storage system can be deployed athwart numerous data-centers. The cloud is responsive of the network topology and will parallel in addition to distribute files transversely the network so that the failure of any one data-center does not limit right to use of data.

D. Users typically assign territories in numerous ways

Computer Rack or Network Switch: This configuration allows an administrator to instruct a storage cloud to distribute files and functionality across systems within a single data center. Data-center: This pattern permits an administrator to notify the cloud to allocate files and functionality between data-centers. User-Based: For storage clouds that span numerous geographies, it is beneficial to inform the system which computers are dedicated to entity user groups. Often this is a parallel configuration to the data center option. Hardware-Based: This configuration permits dissimilar configurations of hardware to be grouped mutually. Region settings can be configured on a machine-by-machine basis: Administrators can decide from some of these use cases otherwise develop hybrid configurations that meet their desires [13].

E. Battles for user's data quickly in the cloud

File spot: The computer servicing a file request places which machines in the cloud hold the requested file using dependable hashing and a circulated hash tables. No single machine holds the total file directory, as it would become a recital blockage or a position of collapse. Lookups are a constant time operation that returns the machines within the system holding a replica of the file. Machine choice: Once the target machines holding the file have been recognized, the requesting machine can choose which machine is finest for retrieving the file. This selection can be made based on factors such as network proximity and machine utilization. File Retrieval: Once the machine is elected, the file can be retrieved by the client. Disaster Recovery: The storage system is entirely distributed; there is no middle position of collapse. Cloud storage can carry on process even when total data-centers have been removed from the system. Cloud storage in addition to controls the merging of numerous data-centers after a logical or physical separation occurs. Availability: Every computer in the cloud system is proficient of serving access to files or administrative requests. Cloud storage is merely able to service a large number of client requests by distributing the work across numerous machines. The system is impervious to the loss of individual or even total racks of machines. Manageable: Administrators are able to revise computer configurations, system configurations, or revise the cloud system itself without taking the files offline.

Management and ease-of-use features are necessary for the formation of a robust cloud storage system: When dealing with hundreds or thousands of machines, management processes must be simplified. Always Available Operation: Any configuration changes made to the system should not take out the ease of use of files. In the occasion that numerous machines want to be taken off-line for revises, the system should have a strategy for keeping files available. Configurable Reliability Settings: Administrators can announce how many copies of a file should be stored in the storage cloud. A cloud-wide setting is time-honored, which may be overridden on a file-by-file basis. Real time computer addition: When the storage system needs additional capacity, the administrator needs to be able to put in machines without affecting the availability of any file. Real time computer is retreating: When it is decided that a computer is no longer required, the administrator needs operations to gracefully take out the computer from processing requests and move its files to other machines within the cloud. Auditing: Important processes, events, and system messages require to be saved. System-Wide Configuration Changes: All configuration changes require propagating across all machines with a single process.

1) *Heterogeneous:* Not all machines in the cloud system want to be constructed from related hardware. The system needs to recognize the attributes of each attached computer and utilize their resources accordingly.

2) *Web Services obtain Web-applications to the next stage:* By using Web services, your application can pass its function or message to the rest of the globe. Web services make use of XML to code along with to decode data, and SOAP to transport it. With Web services, your accounting sectors Windows 2000 server's billing method can connect with your IT supplier's UNIX server.

Web Services have two forms of uses: Reusable application-components. There are things applications needs very often. Web services can offer application-components like: currency translation, climate reports, or even language conversion as services. Connect existing software. Web services can assist to solve the interoperability problem by giving different applications a method to connect their data. With Web services you can exchange data among different applications in addition to different platforms.

III. END USER'S BIG-DATA ON CLOUD COMPUTING

The public cloud is considered the least protected of the three types, with services and resources able to be accessed over the Internet through protocols adopted by the provider [14]. The communications protocols adopted by the provider are not necessarily protected; the choice of using protected or non-protected protocols are up to the providers. The public cloud is also the least expensive of the cloud types, with costs savings in the parts of information technology deployment, management, along with maintenance.

A. Methods with Tools for Mastering Big-data on the cloud

The ultimate application of cloud technology (Fig.2) is as a large-scale data storage, development and processing system, allowing enterprise to master big data. The application of cloud computing technology can assist to get immediate impact from cloud technology while setting a course for mastering the big-data. But the agility of cloud computing has applications ahead of effective use of data. Because all data is now maintained in a centralized system, and can assist develop and implement a centralized security policy that can be easily enforced, allowing precise and well-documented control of susceptible data. In addition, the cloud provides an environment in which to prototype, test, and deploy new applications in a portion of the time and cost of traditional systems [1]. While more data-sets are aggregated, the cloud gains a significant collection of data across an enterprise, apt "the place" on the way to lay data. While each data-set is added, with potentially analyzed through the other data-sets, there is an exponential rise in advantage toward the enterprise, can allow venture among simplified programming furthermore data models, which, combined amid simple access to a ample range of data, results in an report of innovation from crosswise the enterprise in the form of data fusions, data-mining applications, also one-time make use of applications. Simplicity: Scalable systems can hold the biggest datasets with ease, resulting in orders of magnitude decrease in development time and complexity. Scalability: Using numerous computers in place of one allows system capacity to increase simply by adding additional machines. There is no exponential increase the rate. Agility: Since individual machines are programmatically proscribed and usually virtualized, their tasks can be promptly changed, based on changing demands.

B. Basic testing concepts
Tester: A tester refers to a particular party who participates in a test activity. Generally speaking, testers can be human beings, organizations and software systems. *Activity:* There are a variety of test activities including test preparation, test case creation, test implementation, result validation; sufficiency measurement and test report creation, etc. *Context:* Test activities may occur in different software development stages and have various test *reasons*. The concept context defines the contexts of test activities in testing processes along with test methodologies. *Method:* For each test activity, there may be numerous applicable test methods. Method is a part of the capability as well as too an non-compulsory part of test task. Test methods can be classified in a quantity of different ways. For example, test methods can be classified into based on program, requirement, usage and etc. They can also be classified into testing types like structural, fault-based, error-based and etc. Structural testing methods can be further classified into the testing types of control-flow, data-flow and etc. Therefore, test methods are represented as a ladder in the ontology. *Environment:* It is the hardware as well as software configuration in which a test activity is performed.

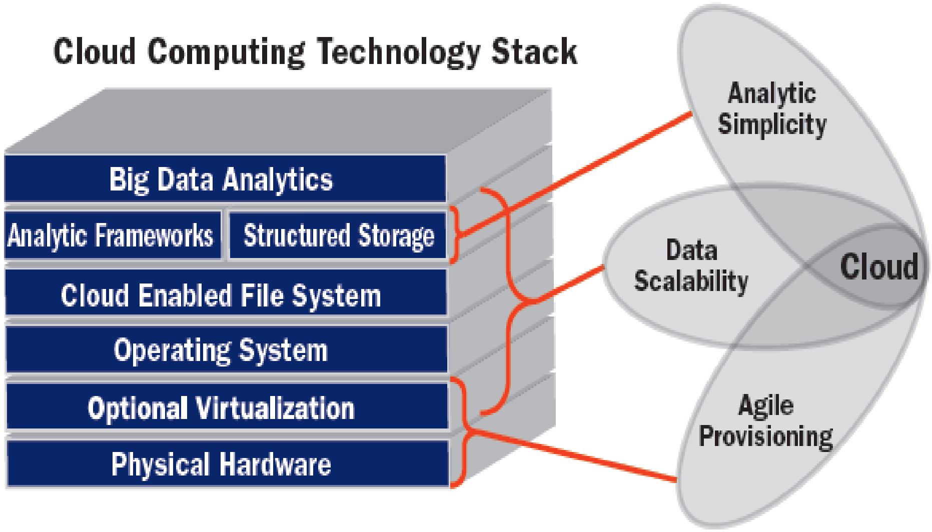

Fig.2: Four layer technology stack with cloud computing

C. Guarantying Testing Quality of SOA Applications
Companies are increasingly moving to services-oriented architectures (SOAs) and deploying Web services within and across their IT infrastructure. These component-based architectures enable powerful applications that exchange information in a simple, standardized manner in addition to permit integration with both internal as well as external application components. However, the hit of these SOA-based application deployments is impacted by the quality as well as performance of the different applications with components by which they interface. Traditional testing solutions focus on testing applications at the user interface level. For SOA based applications however, the user interface might not be vacant for testing until delay in the development series. Moreover various applications or critical interfaces there may not be a traditional user interface at all to test against [15].

IV. OPERATIONAL DISPUTES AND CONCERNS OF BIG-DATA IN THE CLOUD WITH WEB SERVICES TESTING

A. The Needs for big data with cloud computing
Over 70% of responders indicated that they process streaming big-data (Fig.3) – either at large volumes, high velocity, or both. Only 28% said that they do not processing streaming occurrences. Some responders commented that though they are not streaming big-data at current, they are planning to soon, or are looking into fresh ways to gather and process data. They require to process high velocity of big data relates to the need for low latency, such as is obligatory for automatic trading, telecom fraud detection, or e-commerce applications. The need to process huge volumes of data refers to

performance or the need to ensure that the speed of data entry and data processing are aligned so that no backlogs are created, as is obligatory for analytic applications.

B. Advantages of Operational disputes of big-data in the cloud with web services testing

Heavy overheads reduction: The next limit for innovation, competition, along with productivity" monthly thirty plus billion contents are shared through communal media, as well as IT operating expense have increased 6% [16]. Multi-tenancy: The cloud service providers can give gain to a number of clients with low expenditure and the clients feel as if they have their own infrastructure, database, middleware and other services. Multi-tenancy paradigm of cloud environments, policy granularity will be vital to ensure security and compliance. Data integration across cloud platforms will be more of a blockage than application integration, as applications have become more open/standard. Transaction market creation: A huge transaction market can be created through cloud computing and its service management. Speed in big-data processing: enormous volumes of data can be processed at top speed in a cloud environment. Analytics can assist in achieving this task.

C. Web services testing on big-data with cloud computing

Cloud Computing has turn into a scalable services utilization as well as delivery platform in the region of services computing. The technological foundations of cloud computing contain service-oriented architecture (SOA) along with virtualizations of hardware as well as software. The target of cloud computing is to distribute resources among the cloud

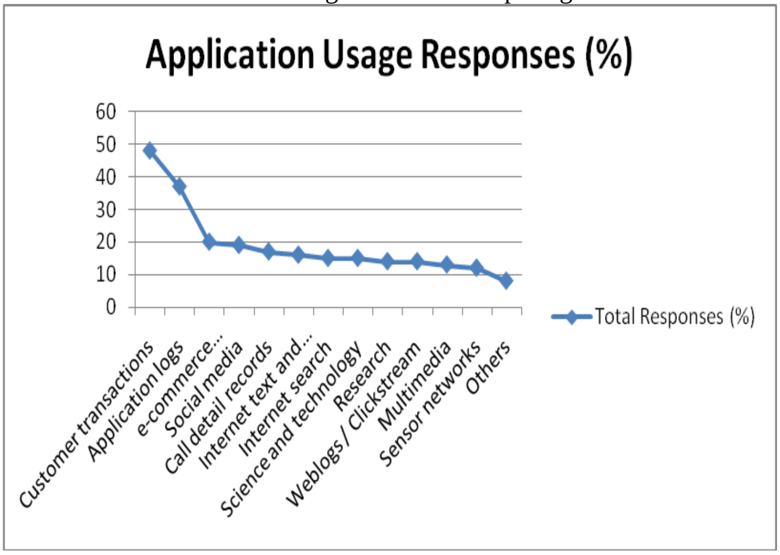

Fig.3: Big data in the cloud responses

service consumers, cloud partners, furthermore cloud vendors in the cloud value chain. The resource sharing at a variety of stages results in a range of cloud offerings such as infrastructure cloud, software cloud, application cloud along with business cloud.

Big-data is a key enabler of exploring transaction insights along with economics of services. Cloud computing furthermore big data are two fast-growing region of information technology. And the force to house as well as analyze increasingly larger sets of data, not including swamping down on-premises systems, creates an association among the two. Big-data in the cloud faces a enormous range that includes analytics, management, tools and more.

Service-Orientated Architecture / Web-services applications are upcoming systems which are enabling businesses to interoperate with are growing at an unprecedented rate. Web-service "end-users" are generally not user web front-ends but added back-end servers. Web-services are exposed to the net like any additional service but can be used on FTP, HTTP, MQ, SMTP, among other transport protocols. It is usually stress free for a tester to journey from one to another technology, except at times it is more difficult to move from one to another methodology. A web-service offers a easy interface for communication for these systems using a usual data transfer method.

D. Ladders in Web-service testing

To wind up what is expected from a Web-service by respect to transaction requirements; To collect with understand the requirements, as well as the data transfer standards; To design test cases keeping transaction requirements in mind, the additional data scenarios you have, healthier the excellence of deliverable. The swindle is to have an automated tool which is able to shorten the testing of web-services like Optimyz, SOAP User Interface and etc (Fig.4).

Performance: Testing web-services routine may be complicated. To stay away from this, following a plain rule of clearly mentioning the thresholds open solves problems. Another type is to recognize the routine requirements in the most perfect manner. For e.g. A good requirement is service has been identified as serving 55,000 to 60,000 concurrent users with 12 secs average response time; A bad requirement is service should serve > 5000 concurrent users, and the response should be speedy. Security: Web-services are wide open in a network.

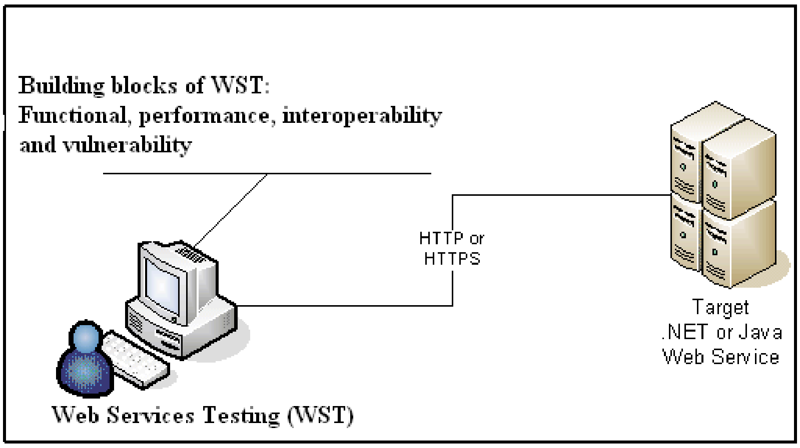

Fig.4: Building blocks of Web-services testing

This factor opens up a host of vulnerabilities, such as Denial-of-Service (DOS) attacks, great volumes of spam data, penetration and etc. There are assured security policies which are enforced through data transfer, along with user tokens otherwise certificates are common sights where data is protected with a code word. Exact test cases meant at directing these policies require to be designed to completely test the Web-service security. Compliance: Compliance testing is required to ensure that the web-services meet certain specified standards; authorize SOAP request / response messages along with Authenticate WSDL definitions.

V. CONCLUSION

Cloud environment is widely used in industry along with research aspects; therefore security is a vital aspect for organizations running on these cloud environments. The integration of big-data as well as cloud computing can provide useful development for both communities. As a result, the problems of big data will only become more famous in the future as solutions are being developed to meet the promising needs. Future development would require cloud and big-data analytic standards. Testing Web-services presents a mixture of new as well as interesting challenges. In particular, test automation will be crucial to a sound as well as proficient Web-service development method, for the assessment of the performance, scalability as well as functionality of Web-services testing in addition to for the approval and acceptance of Web-services developed by application providers.

REFERENCES

[1] Isard, M., Prabhakaran, V., Currey, J., Wieder, U., Talwar, K., And Goldberg, and A. Quincy: Fair scheduling for distributed computing clusters. In Sosp (2009).
[2] Zaharia, M., Borthakur, D., Sarma, J. S., Elmeleegy, K., Shenker, S., And Stoica, I. Delay scheduling: A simple technique for achieving locality and fairness in cluster scheduling. In Eurosys (2010).
[3] Ghodsi, A., Zaharia, M., Hindman, B., Konwinski, A., Shenker, S., And Stoica, I. Dominant resource fairness: Fair allocation of multiple resource types. In USENIX NSDI (2011).
[4] Ananthanarayanan, G., Kandula, S., Greenberg, A., Stoica, I., Lu, Y., Saha, B., And Harris, E. Reining in the outliers in ap-Reduce clusters using Mantri. In USENIX OSDI (2010).
[5] Chowdhury, M., Zaharia, M., Ma, J., Jordan, M. I., And Stoica, I. Managing data transfers in computer clusters with orchestra. In ACM Sigcomm (2011).
[6] Shieh, A., Kandula, S., Greenberg, A., Kim, C., and Saha, B. Sharing the data center network. In USENIX NSDI (2011).
[7] Web Services Conceptual Architecture WSCA 1.0, IBM Software Group, May 2001.
[8] Web Services Architecture[s]. W3C Working Draft, at http://www.w3.org/TR/ws-arch/.
[9] B. De, "Web Services - Challenges and Solutions", WIPRO white paper, 2003, http://www.wipro.com.
[10] Al-Fares, M., Radhakrishnan, S., Raghavan, B., Huang, N., and Vahdat, A. Hedera: Dynamic Flow Scheduling for Data Center Networks. In USENIX NSDI (2010).
[11] Wang, G., Andersen, D. G., Kaminsky, M., Papagiannaki, K., NG, T. E., Kozuch, M., and Ryan, M. c-Through: Part-time Optics in Data Centers. In ACM Sigcomm (2010).
[12] Farrington, N., Porter, G., Radhakrishnan, S., Bazzaz, H. H., Subramanya, V., Fainman, Y., Papen, G., and Vahdat, A. Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers. In ACM Sigcomm (2010).
[13] Chen, Y., Alspaugh, S., and Katz, R. Interactive analytical processing in big data systems: A cross-industry study of Map Reduce workloads. In VLDB (2012).
[14] Phanishayee, A., Krevat, E., Vasudevan, V., Andersen, D. G., Ganger, G. R., Gibson, G. A., and Seshan, S. Measurement and analysis of TCP throughput collapse in cluster based storage systems. In FAST (2008).
[15] H. Zhu and Q. Huo, "Developing A Software Testing Ontology in UML for A Software Growth Environment of Web-Based Applications," *Software Evolution with UML and XML*, H. Yang, (ed.), IDEA Group Inc. pp263-295, 2005.

[16] Hadoop. http://hadoop.apache.org/.

[17] H. Zhu, Q. Huo, and S. Greenwood, “A Multi-Agent Software Environment for Testing Web-based Applications,” *Proceedings COMPSAC’03*, pp.210-215, Nov. 2003.

[18] W3C: Simple object Access Protocol (SOAP) 1.1, W3C Note 08 May 2000, http://www.w3.org/TR/SOAP.

[19] R. Wang and N. Huang, “Requirement model-based mutation testing for web service,” in NWESP ’08:Proceedings of the 2008 4th International Conference on Next Generation Web Services Practices, pp. 71–76, Seoul, Korea, Oct. 2008, IEEE Computer Society.

[20] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, June). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_Innovation.

[21] Sliwa, C. (2011, June 16). Scale-out NAS, object storage, cloud gateways replacing traditional NAS. Retrieved from http://searchstorage.techtarget.com/feature/Scale-out-NAS-objectstorage-cloud-gateways-replacing-file-storage.

