Internet Data Center

Definition and History

A Data center is a facility, which houses thousands of computing systems to run IT enabled business services without any interruption. The infrastructure to store, process and analyze information and to provide services through internet defined by business needs, minimizing disruptions and obstacles related to information systems in the process, defines an Internet Data center. As IT operations scaled globally, the importance of Internet Data center has become grown manifold. A data center is now an integral part of enterprise organizations, beyond reasonable doubt. Enterprises need to handle astronomical amount of data generated every day. This is not a recent phenomena but in an increasingly competitive and open market, most enterprises would have to use the data and create new platforms and solutions. These, in turn, keep the enterprises in healthy financial shape and help maintain the customer base. Efficient customer facing analytics solutions have emerged in many enterprises and this is predominantly, due to the leverage data centers offer. The traditional data center, also known as a “silo-ed” data center, is dependent heavily on physical servers and hardware in stark contrast to Internet Data Centers. The journey of traditional data centers began from mainframe computers and the growth was triggered by the dot.com bubble in the late 1990 and early 2000. Reliability and performance notwithstanding, the expansion of a traditional data center is restricted by physical limitations and constrained by the time to deployment of new applications . The issues in space, scalability and portability led the birth of virtual data centers. Powered by Virtual technology, virtual/internet data centers create a resource pool of computing devices, which can be

reallocated based on demand. The trend was first observed in 2003 and by the end of 2011, 72% of enterprise organizations converted one fourth of their traditional data centers to virtual ones. The global dependence on internet data centers has grown phenomenally in recent times. Demand for storage and maintenance of data is not limited to facilitators of information technology only, but has spread to all forms of businesses and service providers, private, public and individual.

Essential Facts about Data Centers

A Data Center can't survive without Power, Cooling systems and Network Connectivity. Electricity is supplied to large data centers either from grids or on-site generators. Cooling systems are required to mitigate the heat dissipated by power supplies to the computing systems. Chillers are used to supply cold water, necessary for air conditioning systems. Facebook deployed one of its largest data centers in Alaska. Typically, the information and computing systems are organized in computer racks. A rack is an Electronic Industries Association enclosure, which is 2 meters high, 0.61 meters wide and 0.76 meters deep. A standard rack accommodates 40-42 computing units and dense rack configuration servers (Blade rack) may accommodate 200 computing units. The heat dissipated by a standard rack is 10KW and Blade rack will dissipate heat up to 30KW. Typically, a data center containing 2000 racks may require 20MW power. Uptime Institute has defined four tiers of data center based on destination availability. Tier 1. Tier 2, Tier3 and Tier 4 guarantee availability of 99.671%, 99.741%, 99.982% and 99.995% of data respectively. Of late, data centers are effectively being utilized in cloud computing, known as Cloud Data Centers. Cloud is an example of off-premises computing

where services are now catered through data centers. Cloud service providers also own data centers, which are located in different geographical locations for the provisioning of uninterrupted services in case of outage and unpredictable incidents. IaaS (Infrastructure as a service), which provides facilities namely virtual machines, storage and load balancing maintains a large pool of resources in data centers.

The Flip side: Cost of Data Centers

The major costs incurred by data centers are Infrastructure, Power and Cooling. Rising energy cost and responsibility towards the environment has put data centers under tremendous pressure to improve energy consumption. According to a survey, data center consumed 1.3% of total global energy usage in 2010. A minor improvement in energy consumption will have significant impact on cost, saving millions of tons of carbon emission. The most well known metric, used to measure the energy efficiency of a data center is power usage effectiveness (PUE). An ideal scenario demands the value of PUE to be 1, indicating no overhead energy consumption apart from IT equipment. Google's PUE performance has improved from 1.21 in 2008 to 1.12 in 2013 due to the adoption of best practices and natural progression down the learning curve. The cost of power to run data centers is expected to dominate among all other costs. Some of the recent technological advances such as dynamic smart cooling techniques, equipped with temperature-aware cooling algorithm and using a scale processor and system power, have been adopted to reduce cost. In the year of 2007, Google has spent $2.4 billion on its data centers, the expenditure ramping up to $11 billion in 2015.

Data Center and the future of Internet: Is it good Economics?

The economics of the data centers don't enjoy the attention it deserves. IT organizations need to invest heavily to set up their own data centers. Small firms rent space rather than building a new facility. The rising cost of power supply, the crunch in storage space owing to real estate bubbles, the difficulty in acquiring land for industrial use in various countries, etc., translate into important trade-offs for data centers. The growth of business and competition leads to lower per unit prices, but the costs have escalated across the board. The big firms seem to cope with rising cost because the market for data center operations is fairly concentrated. Indeed, in the recent years, a lot of investments have been made in data centers in order to support cloud computing by large organizations. However, casual empiricism suggests that this industry usually harbors a large number of firms and therefore deviations from least cost combinations of inputs owing to exogenous shocks could be potentially disastrous for many companies affecting the scale of operations. In many cases, cost reducing innovations and potential for flexibility are rather important, but these are often outcomes of sustained and costly research and development activities at the firm level. The larger firms are more likely to engage in such activities. The issue is particularly compelling and throws open questions worth pondering. Is the data center market competitive enough? Will the small firms wither the onslaught?

Current and Future Research

The current decade is witnessing an implosion in academic and industry sponsored research in data center. One of the hot topics is the fusion of big data and cloud data center. The emergence of cloud computing and high performance platforms such as Hadoop and Scala lay the foundation for big data analytics exploiting the leverage provided by internet data centers. Internet data centers rely on distributed computing, prone to security threats. Data security is a concern and Big data security, in particular is an area where a lot of research activity is taking place. Researchers are increasingly focusing on optimization of the various cost components, which will consequently improve the revenue of the organization and possibly throw some light on the sustainability of small and medium enterprises in this business. Significant research activity has been reported in literature regarding load balancing of incoming requests in data centers. Perhaps, data center economics and management are the areas to be closely watched in near future.

Snehanshu Saha and Jyotirmoy Sarkar

See also Blogs as a Dot-Com Bubble; Cloud Computing; History of the Internet; See Internet Origins and History; Internet Origins and History; Internet of Things; Smart Energy Systems; Smart Grids; TCP/IP; Mobile Internet.

Further Readings

Patel, C. D. ,& Shah, J. A. (2005). Cost Model for Planning, Development and Operation of a Data Center. Retrieved February 14, 2017; from http://hpl.hpl.hp.com/techreports/2005/HPL-2005-107R1.pdf

Hamilton, J. (2008). Cost of Power in Large-Scale Data Centers [WWW 1190 document]. Retrieved March 9,2015; from http://perspectives.mvdirona.com/2008/11/cost-of-power-1191in-large-scale-data-centers/.

Koomey, J. (2011). Growth in Data center electricity use 2005 to 2010. Analytics Press, Oakland, CA.

Saha, S., Sarkar. J., Anand M. N., Dwivedi, A., Dwivedi, N., Roy, R., & Rao, S. (2016). A novel revenue optimization model to address the operation and maintenance cost of a data center. Journal of Cloud Computing, (Springer), 5,1; 1-46.

Sarkar, J., Goswami, B., Saha, S., & Kar, S. (2016). CDSFA Stochastic Frontier Analysis Approach to Revenue Modeling in Large Cloud Data Centers. Retrieved February 14, 2017; from https://arxiv.org/abs/1610.00624