

Chapter 12

Examining Data Lake Design Principle for Cloud Computing Technology and IoT

Deepak Saini

Publicis Sapient, India

Jasmine Saini

Jaypee Institute of Information Technology, India

ABSTRACT

In the Cloud-based IoT systems, the major issue is handling the data because IoT will deliver an abundance of data to the Cloud for computing. In this situation, the cloud servers will compute the big data and try to identify the relevant data and give decisions accordingly. In the world of big data, it is a herculean task to manage inflow, storage, and exploration of millions of data files and the volume of information coming from multiple systems. The growth of this information calls for good design principles so that it can leverage the different big data tools available in the market today. From the information consumption standpoint, business users are exploring new insights from the big data that can uncover potential business value. Data lake is a technology framework that helps to solve this big data challenge.

1. HISTORY AND EVOLUTION

In this section, the foundational concepts about Data Lake are covered. This section will start with why and how the necessity of Data Lake arose, the various data lake related terminologies prevalent in industry, justify the need of Data Lake. Some design principles for Data Lake will also be discussed.

DOI: 10.4018/978-1-5225-3445-7.ch012

In order to efficiently store, organize, process, and present data, architects came up with a number of design approaches. These data management approaches are good for IT, however, these are not agile and easily understandable by business. Also its turnaround time and cost is higher than what business expected. This created a barrier between IT and business.

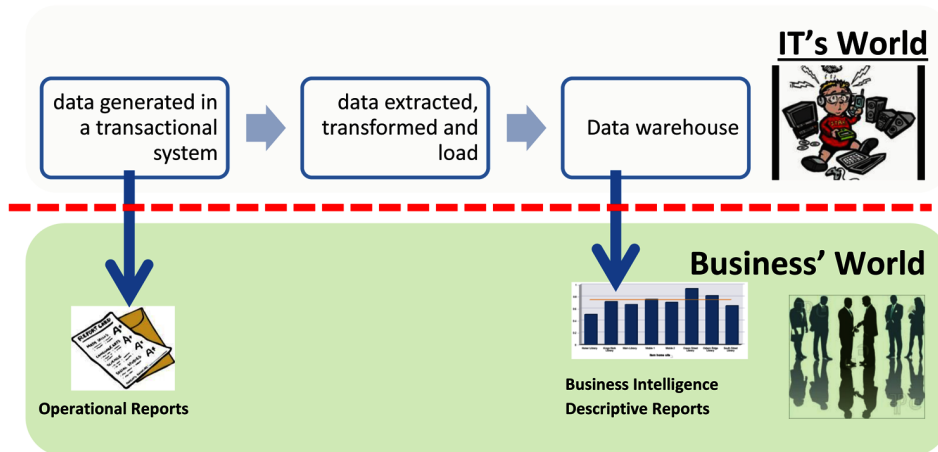
Data architecture and data governance created by IT was rigid, as they constrained by high cost of storage, slow processing power of compute, whereas business needed agility and interactive insights fast. The time to change or implement a new request for a business insight was too slow. There was burning need to overcome this mismatch between IT and business, and increase the agility in data management solutions. Hence, it became necessity that both business and IT collaborate to manage the data. Figure 1 represents the traditional design approach for processing data for reporting.

In new era of data management, the traditional concepts of Online Transactional Processing (OLTP) systems, Online Analytical Processing (OLAP) systems, Data warehouse (DWH), Relational Database Management (RDBMS), Business Intelligence (BI) is augmented, but not replaced, by new concepts like Big Data Processing, Data Lake, No SQL, Predictive and Prescriptive Analytics, Data Science.

These new disciplines introduced whole new approach for data management. The approaches promote agility through flexible schema design, store all and discard none philosophy, massive parallel processing, distributed computation, etc. These new approaches are now possible because –

- Storage cost has become cheaper
- Computation of data is distributed in multiple machines.

Figure 1. Traditional design approach for processing data for reporting



21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/examining-data-lake-design-principle-for-cloud-computing-technology-and-iot/191841

Related Content

A Review of Intrusion Detection Systems in Cloud Computing

Chiba Zouhair, Noredine Abghour, Khalid Moussaid, Amina El Omriand Mohamed Rida (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 54-83).

www.irma-international.org/chapter/a-review-of-intrusion-detection-systems-in-cloud-computing/224567

Resource Management in Vehicular Cloud Computing

Sawsan Elfatih Ibrahim, Rashid A. Abd Elhaleem Saeedand Amitava Mukherjee (2018). *Vehicular Cloud Computing for Traffic Management and Systems* (pp. 75-97).

www.irma-international.org/chapter/resource-management-in-vehicular-cloud-computing/206611

Social Implications of Big Data and Fog Computing

Jeremy Horne (2018). *International Journal of Fog Computing* (pp. 1-50).

www.irma-international.org/article/social-implications-of-big-data-and-fog-computing/210565

Fog Computing Quality of Experience: Review and Open Challenges

William Tichaona Vambe (2023). *International Journal of Fog Computing* (pp. 1-16).

www.irma-international.org/article/fog-computing-quality-of-experience/317110

Enabling Open and Collaborative Public Service Advertising through Cloud Technologies

Daniel Pop, Alejandro Echeverria, Dana Petcuand Gloria Conesa (2016). *Cloud Computing Technologies for Connected Government* (pp. 269-290).

www.irma-international.org/chapter/enabling-open-and-collaborative-public-service-advertising-through-cloud-technologies/136882