IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

Analysis of Data Virtualization

P. Srinivasulu[1], C. Yosepu[2], B. Subbarayudu[3], A. Kavitha[4]

Assistant Professor, Dept of IT, St. Martin's Engineering College, Hyderabad, India[1, 3]

Assistant Professor, Dept of CSE, St. Martin's Engineering College, Hyderabad, India[2, 4]

Abstract: Virtualization provides several benefits –greater efficiency in CPU utilization, greener IT with less power consumption, better organization through essential environment control, more availability, reduced project timelines by eliminating hardware procurement, enhanced disaster improvement capability, more central control of the desktop, and improved outsourcing services. Many blogs and articles are devoted to data virtualization, at many events data virtualization has been discussed the issue is explained on numerous webinars and webcasts, it's on the radar of nearly all analyst organizations, all types of organizations are using the technology nowadays, and the products have matured sufficiently to handle large and complex data environments. One can simply state that even though the term data virtualization is not as popular as the terms SQL, data warehouse or big data, data virtualization has been accepted by the market. More and more organizations are deploying the technology to simplify right to use to their labyrinth of data sources.

Keywords: Virtualization, data warehouse, big data, organization

I. INTRODUCTION

Data virtualization allows an organization to compose its activity data easily available to business users [4]. From a further scientific standpoint, data virtualization makes all the activity data that has been dispersed over a multitude of IT systems look like one reasonable database—still if that data is covered deeply in IT systems[1]. The effect of using data virtualization is that organizations can enhance the return on their investment in data processing [2]. Virtualization is the formation of a virtual (rather than actual) description of object, such as an operating system, a server, a storage device or network resources [3].

Data Processing:

The collection and exploitation of items of data to produce meaningful information. In this sense it can be considered a split of information processing, the modify (processing) of information in any manner detectable by an observer. The phrase is often used more particularly in the context of an industry or other organization to refer to the class of commercial data processing applications.

Data virtualization is relatively young. Exactly when the term was first coined is not completely coherent. It looks as if Eric Broughton used the term first in a paper published in 2005.Although not as admired as terms such as big data and cloud, for the last five years data virtualization has moved slowly into the spotlights.

The history of data virtualization is powerfully related with data federation, which has been around much longer. (For additional information about that relationship, see my object, Clearly Defining Data Virtualization, Data Federation and Data Integration.) Data federation mechanism combining a various set of autonomous data stores to form one large data store. In standard, this is what data virtualization does as well. But this is where data federation stops and data virtualization continues. Subsequently to data federation technology, current data virtualization products also support purification technology, data profiling and data modeling capabilities, impact and lineage analysis and so on. Some products that ongoing out as pure data federation products evolved into data virtualization products.

Data virtualization, attempts to execute data cleansing, data transformation and data correlation as data moves out from production systems thus avoid any intermediate storage. This is opposed to Data warehouse approach which physically changes data in every stage and loads it in to some data store. Typical Data virtualization platform requires:

- Ability to determine data stored in data sources
- Ability to recover data from different data sources
- Ability to define views or virtual tables
- Ability to optimize federated query
- Ability store data
- Fine grained security

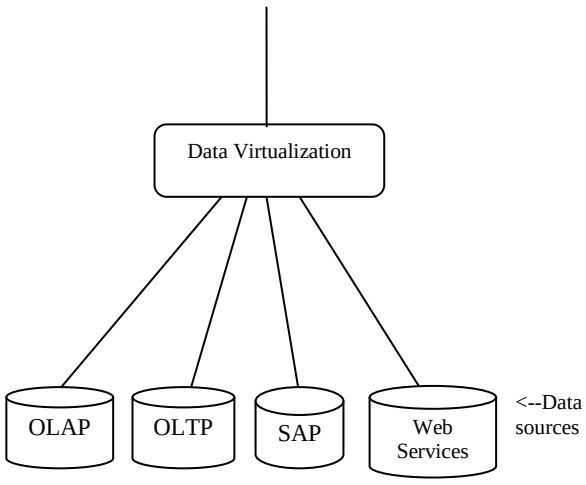

Figure 1: Over View of Data Virtualization

DOI 10.17148/IJARCCE.2015.41282

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

The above diagram depicts usage of Data virtualization platform for integrating data from databases, SAP and web services. Each phase of data analysis gets defined via virtual tables. When analysis needs to be done Data virtualization compiles all of the definitions of virtual tables within to a single SQL, which is then compiled, optimized and executed.
Virtual Tables or Views are the outcome set of a stored query which can be used just like a regular database table. The table schema and table contents are defined through SQL. The table is considered virtual because table contents are not actually stored. Data for the virtual table is brought in from underlying database tables when query defining virtual table is executed.

Data Federation
In almost all cases, if the phrase federation is used, it refers to combining autonomously working objects. For example, states can be federated to structure one country. If we concern this common explanation to data federation, it means combining independent data stores to form one large data store. Hence, we propose the following definition:

Data federation is a structure of data virtualization where the data stored in a heterogeneous set of autonomous data stores is made available to data consumers as one integrated data store by using on-demand data integration.
This description is based on the following concepts:

Data virtualization
Data federation is a form of data virtualization. Reminder that not all forms of data virtualization involve data federation. For example, if an organization needs to virtualized the database of one function, no require exists for data federation. But data federation always results in data virtualization.

Various set of data stores
Data federation should make it possible to bring data together from data stores using special storage structures, special access languages, and different APIs. An application with data federation should be able to access special types of database servers and files with various formats; it should be able to combine the data from all those data sources; it should recommend features for transforming the information; and it should allow the applications and tools to entrance the data during different APIs and languages.

Independent data stores
Data stores accessed by data federation are able to operate independently; in other words, they can be used external the range of data federation.

One integrated data store
Regardless of how and where data is stored, it should be offered as one integrated data set. This implies that data federation involves transformation, cleansing, and maybe even enrichment of data.

On-demand integration
This refers to when the information from a various set of data stores is integrated. Among data federation, integration takes rest on the fly, and not in batch. When the data consumers request for data, only then data is accessed and integrated. So the information is not stored in an integrated way, but remains in its original location and format.

II. THE PRESENT MARKET FOR DATA VIRTUALIZATION

The most current study of the advertise for data virtualization was performed by Wayne Ackerson of Tech Target in April 2013 (observe Data Virtualization: Perceptions and Market Trends). This study shows to facilitate 35% of the respondents have invested currency in data virtualization, 27% of the respondents have incompletely deployed the software and 18% have it completely deployed. Moreover, almost one-third of the organizations have data virtualization within consideration.

These numbers are equivalent to the ones upcoming from a TDWI (The Data Warehousing Institute) study. This study indicates that 19% of the organizations have data virtualization presently in use and 31% have plans to implement data virtualization. In July 2012, Ted Friedman of Gartner indicated that around 27% of the respondents of a research indicated that they were actively concerned in or had strategy for deployment of federated or virtualized views of data. Ventana study indicated in an April 2012 study that data virtualization is an advancing main concern in information management: 12% have accomplished data virtualization projects, 11% have initiated projects and 20% have designed a project in which data virtualization will be used. Lastly, in a 2012 revision, Forrester Research predicted that the total software revenues (licenses, preservation and services) designed for data virtualization would grow to $8 billion by 2014.

These records are very promising, but plainly the data virtualization market is not growing explosively. It's growing as so several enterprise software products grow: slowly and steadily. In addition, we have to remind ourselves that it was only throughout the last five years that data virtualization vendors have started to push and promote data virtualization heavily, and that period coincides with a global market under stress. As is generally known, in a poor economy, organizations invest less in innovative technologies, even if that technology may solve some of the problems and lower the entire cost of ownership (TCO).Noteworthy is that, based on conversations with dominant data virtualization vendors, Europe is extensively behind the USA with respect to adopting data virtualization.

III.HOW ORGANIZATIONS ARE USING DATA VIRTUALIZATION

While some of the data virtualization products were initially considered to maintain ESB/SOA type systems, at present in business intelligence environments many organizations are used to data virtualization. This was also shown by Wayne Ackerson's in the past mentioned study. These were the use cases for data virtualization:

- Enhance the data warehouse (77%)
- Prototyping (45%)
- Visualizing concurrent data within an existing application (39%)
- Drill into complete data in one more system, such as a data warehouse (30%)
- Querying non-relational data sources (25%)
- Creating an activity view of numerous data warehouses (24%)
- Querying peripheral data (19%)
- Supporting ETL processing (18%)
- Querying peripheral data (16%)
- Delivering a 360-degree view of consumers (15%)

In several cases, organizations are involved to data virtualization as of its agility: the speed with which data sources can be integrated and the velocity with which integrated data becomes accessible to end users. Ackerson's study confirmed this: 66% of the respondents were attracted in data virtualization because of agility. In the long run, the fast growing interest in self-service BI tools, such as QlikView, Spitfire and Tableau will boost the adoption of data virtualization. The cause is that self-service BI tools can only carry a certain level of agility. The moment users ask for a change of a data formation in a data mart or data warehouse, the IT department has to be concerned. Increasing a BI system using self-service BI tools together with data virtualization servers leads to a much more alert result. In other words, the same stage of quickness currently presented by self-service BI tools for reporting and analytics can be delivered by data virtualization servers for the data storage aspects of BI environments. It is also significant that those not too recognizable with the technology assume it can be deployed in small environments simply. According to Ackerson's study this is not true. This study shows that a majority of the organizations (59%) that contain deployed data virtualization have implemented it on an enterprise scale, only one-quarter (25%) have deployed it on industry unit level and just 14% have deployed it at the departmental level for one or more departments. In other words, the popular deploy data virtualization on enterprise scale, which could only mean it's suitable for large scale environments.

IV. FEATURE EXTRACTION

It is essential that the data virtualization products maintain to develop in the following three areas: The vendors have to maintain their research into improving presentation for all kinds of queries and on all kinds of data sources, including NoSQL systems. Moreover, their caching mechanisms have to be strengthened. Here, support for in-memory database servers could be useful. Furthermore, now that more and more data is moving to the cloud, data virtualization products have to progress the efficiency of moving data in, to and within the cloud.

The products have to be extended to support the complete system development cycle. The design modules of current data virtualization servers should allow designers and analysts to enter other further business-related specifications, such as business glossaries, data models and taxonomies. Data virtualization servers should maintain more and more features currently supported by tools such as data modelling, master data management and business illustration tools. Data virtualization servers should be able to support the whole method of information management, as well as information modelling, data governance, and logical database design, and not presently the implementation phase. For more information see my book, Data Virtualization for Business Intelligence Systems.

NoSQL systems are very influential with respect to storage and processing of massive amounts of data. However, in BI environments the majority of the tools in use don't know how to handle data not stored in a relational way. This means that all that expensive (big) data stored in NoSQL systems is not available to everyone in the organization. To open up all that data, SQL interfaces are very significant and useful. A SQL interface to a NoSQL system makes the data presented to a larger user area and thus increases the potential business value of it. Lately, more and more SQL interfaces are fetching available for NoSQL. Unfortunately, these interfaces are still very young and most of them offer contact to one NoSQL system only—they don't federate data from multiple systems. This is a potential area in which data virtualization products can play an important position. They have mature SQL optimizers capable of handling large amounts of data, they have been considered to federate data and they know how to handle non-relational data sources. They should be able to win a extensive section of this market. Thus, the increasing NoSQL market may create a new market for data virtualization products.

V.CONCLUSION

Data virtualization is an acquire technology whose adoption is accelerating steadily. Many IT specialists know what the technology has to suggest and how it makes data architecture more flexible. Studies show that 30-35% of the organizations study, invest and/or deploy the technology at present. The BI market will maintain to push the deployment of data virtualization, but the fast adoption of NoSQL systems will also enlarge the insist for data virtualization. In short, data virtualization is not exposure; it's a reality. Most vendors can show an inspiring list of organizations deploying the technology today. This is supported by the studies done by distinguished analyst organizations.

REFERENCES

[1] E. Broughton, Periscope: Access to Enterprise Data, TUSC, 2005.
[2] D. Stodder, Achieving Greater Agility with Business Intelligence, TDWI, January 2013; see http://tdwi.org/research/2013/01/tdwi-best-practices-report-achieving-greater-agility-with-business-intelligence.aspx
[3] Ventana Research, The Evolution of Information Management, April 2012; seehttp://www.ventanaresearch.com/infomgt/
[4] N. Yuhanna and M. Gilpin, The Forrester Wave: Data Virtualization Q1 2012, Forrester Research, January 2012; seehttp://www.informatica.com/Images/1888_forrester-wave-data-virtualization_ar.pdf

BIOGRAPHIES

P Srinivasulu is an Assistant Professor in the Department of Information Technology, St. Martin's Engineering College, JNT University, Hyderabad. He received Master of Technology (Computer Science and Engineering) degree in 2014 from AITS, JNT University, Anathapur, AP, India. His research interests are, Computer Networks, Big Data and Mobile Computing.

C. Yosepu is an Assistant Professor in the Department of Computer Science and Engineering, St. Martin's Engineering College, JNT University, Hyderabad. He received Master of Technology (Computer Science and Engineering) degree in 2013 from SSIT, JNT University, Hyderabad, Telanagana, India. His research interests are Design and Analysis of Algorithms, Big Data; Data ware Housing and Data Mining, Computer Networks

Bathala Subbarayudu is an Assistant Professor in the Department of Information Technology, St. Martin's Engineering College, JNT University, Hyderabad. He received Master of Technology (Computer Science and Engineering) degree in 2014 from SRTS, JNT University, Anathapur, AP, India. His research interests are Network Security Computer Networks and Big Data.

A. Kavitha is an Assistant Professor in the Department of Computer Science and Engineering, St. Martin's Engineering College, JNT University, Hyderabad. She received Master of Technology (Computer Science and Engineering) degree in 2014 from SMEC, JNT University, Hyderabad, Telanagana, India. Her research interests are, Information Security, Network Security and Mobile Computing.

DOI 10.17148/IJARCCE.2015.41282