A Comparative Study on Typical Data Integration Methods and Technology

[1]Abirami.K , [2]Dr.Gopinath Ganapathy

[1]Research Scholar, Bharathidasan University ,[2]Professor, Bharathidasan University

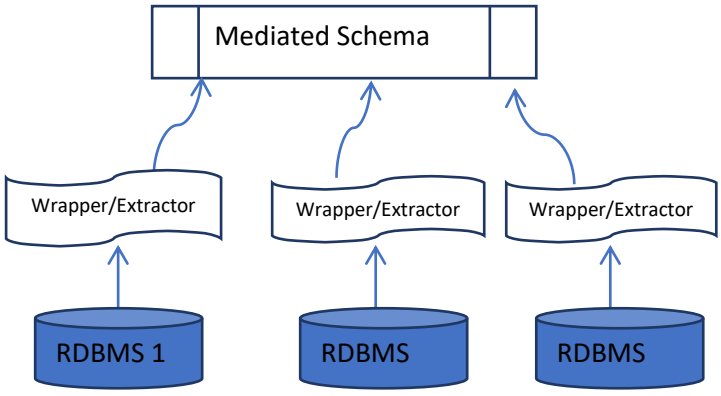
Abstract:

Data integration is the process of pouring several spouts of data into a singlerepository. Despite of various integration methods and techniques are evolved to adjoin the data from diverse data source, choosing the most effective integration tactics, and overseeing the integrity, structure and schema of end data is challenging. This paper presents a comparative study on existing data integration methods and technology.

Keywords: Data Integration, Data virtualizations, Schema

Introduction:

Data integration system access to a set of data sources and have the system automatically configure itself so that it can correctly and efficiently answer queries that span multiple sources.This aims to reduce the effort required to integrate a various data sources and let the schema matches to the new datasource



The wrappers or loaders request and parse data from the sources. The mediated schema provides an abstract view of data and handles queries over this.

Data Integration Techniques

ETL (Extract, Transform, Load)

ETL is the known technology for integration of data as a batch job.it retrieves the data from a source and transforms the data to the as per the destination source format finally load the data

into the destination. ETL is also suited for change data capture (CDC-transfer updated data to the target) .Huge amount of data transferred through bulk movement technique.

Enterprise Data Replication (EDR)

EDR propagation techniques make the dataset to interlink with heterogenous dataset by following same schema. The process of heterogenous data integration becomes complex and moreover data replicated at a regular interval time should meet the needs of the organizations

EDR and ETL handles huge set of data movement whereas EDR does not specific to the transformation and manipulation of data.

Another architecture called Enterprise Application Integration (EAI) with Change Data Capture (CDC) facilitates enterprise data transfer and captures the data changes using logs and base triggers between data source and any destinations.

Enterprise Information Integration (EII)

EII creates a virtual layer for the data resides in the storage. This layer protects the data from the incoming various formats of data from the disparate sources. It sees the information of the organization in a unified way and maintain as a single source.EII has the features to correlate and aggregate the data using a single query

With single data layer EII means the real time-oriented system integration.

Data Virtualization:

This technique creates a virtual layer wherein queries are processed agreeing to the reporting or business intelligence system. From the connecting applications perspective it appears as a single repository but in real data is retrieved and processed from the various sources. The virtual layer handles all the complications in data transformations so as to the source system formats remain unchanged.

Data virtualization also termed as “Logical data warehouse” by means of ease access to the unstructured data types.

XML and Data Integration

XML is a standard to define integrated functionalities.

XML is an enabling standard that companies leverage to define and implement open, integrated business functions.

Since XML uses simple terminologies and expresses the business entities it can be integrated closely using few key word managements. Due to the powerful data integration capabilities of an XML an organization can describe the business assets and ledgers thoughtful.

Big Data Integration:

To increase the scalability and high-performance rate all components of big data ecosystem follows different approaches in integration. When Real-time or batch processing system uses the big data integration ETL sequences are reordered to turn like an ELT ,so the data is retrieved and stored into a globalized distributed systems and then deliver the data. Following core techniques used for Big Data Integration.

- Schema Mapping
- Record Linkage
- Data Fusion

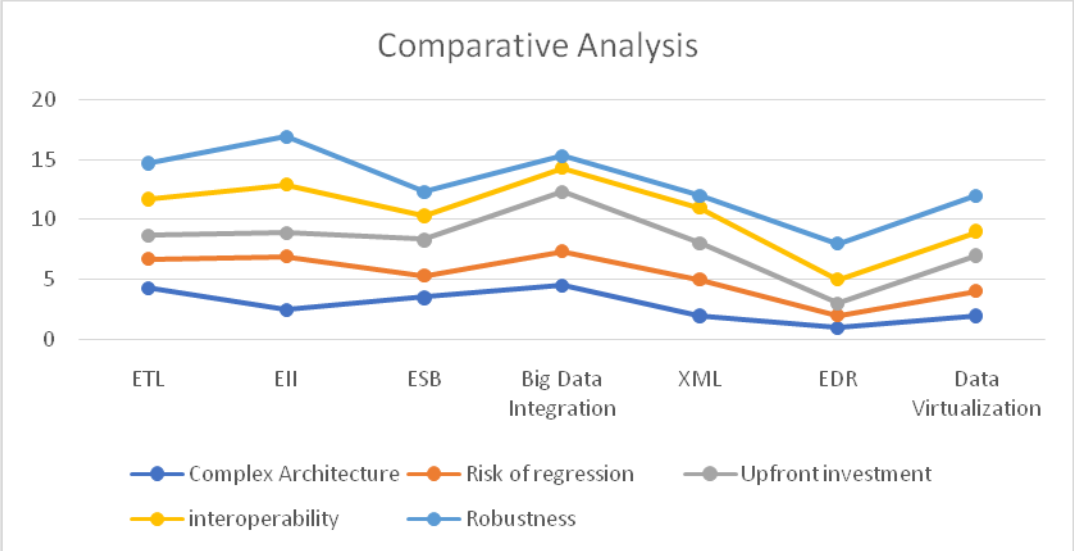
Emergence of bigdata suggested an architecture where data moving is handled by the middleware technology. Middleware decouple from the various data sources and data stores.

ESB and Data integration

An enterprise service bus facilitates an API for data integration**.** It's an architecture provides a structure to integrate enterprise data. Based on the principle of service-oriented architecture ESB control the data much like a router.

ESB adopts infrastructure as a service architecture using which it handles all the data interpretation, routing the information. It restructures the data according to the application requirement. Rule-based engine enforce the data content to adhere the rule set.

Comparative Analysis :



Conclusion:

Thus, as the need of data integration across multipledomains, platform, system increased we emerge to take up the effective integration methods. To get hold of any integration method Data Vulnerability is the major factor to consider.Data should not be leaked, transformed, rather it must be reformed according to the destination.Though various strategies in integration choosing a better solution is based on the data model.

References:

[1] M. R. Wigan and R. Clarke, —Big Data's Big Unintended Consequences‖, IEEE Computer Society, , vol. 46, no. 6, (2013), pp. 46-53.

[2] Abdul Raheem Syed, Kumar Gillela, Dr. C. Venugopal, "The Future Revolution On Big Data", In International Journal of Advanced Research in Computer and Communication Engineering, 2013.Volume: 2 (Issue:6) , Page No. 2446- 2451.

[3] Proyag Pal,Triparna Mukherjee , Dr. Asoke Nath-Challenges in Data Science: A Comprehensive Study on Application and Future Trends, August 2015,Volume 3, Issue 8.

[4]https://importoioweb.staging.wpengine.com/post/best-big-data-tools-use/.

[5]https://towardsdatascience.com/the-10-statistical-techniques-data-scientists-need-to-master-1ef6dbd531f7.

[6]http://www.datascientists.net/what-is-data-science.