International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2064139

Data Integration Techniques for Cross-Platform Analytics

Sai Krishna Shiramshetty

Independent Researcher, USA

Article Info

Volume 6, Issue 4

Page Number: 593-599

Publication Issue :

July-August-2020

Article History

Received: June 29, 2020

Accepted: July 11, 2020

Published: July 20, 2020

ABSTRACT

This paper highlights various data integration techniques that I applied when implementing cross Platform analysis including ETL integration techniques, API integration and real-time integration techniques. This looks at examples from Netflix, Amazon and Uber how these enable organisations to aggregate information from various sources for decisions, customisation and business operations. The consequence of concerning working discussed in this paper is an evaluation of the potentials of effective data integration and presenting it as a crucial factor towards the realization of competitiveness and innovation.

Keywords : Integration of data, Cross Platform Analysis, Data Extraction, Application Programming Interface Integration, Real-time data consumption

I. Introduction

The key to cross-platform analytics is important for any organization that wants to gain insights from data gathered from different platforms. ETL, API integrations and real-time data processing are key methods of data integration required in fusing this disparate information for analysis. Such methods help in the right decision-making process, organization activities, and providing customers with better experiences. The combination of data from multiple platforms is a distinct advantage as it allows organizations to meet market needs and customer requirements within a short time.

II. Literature Review

In the field of multiscreen measurement, data aggregation emerges as one of the significant issues that must be solved for an organization to derive insights from multiple systems and sources. Based on the platform and technology that a business organization uses, then there is need to integrate and consolidate data from the various platforms for analysis. Data integration methods allow combining data from different sources into a format that can be easily analyzed and managed, so businesses profit from knowledge produced from otherwise isolated data sets.

This paper discusses data integration and methods required in the cross-platform analytics to increase the speed and efficiency of organizational decision-making processes. In reviewing literature on data integration techniques, it was observed that a vast variety of methodologies and approaches to data integration exist during the last decade, all having their own advantages and disadvantages.

Of all the procedures of managing the compatibility of data between applications, the ETL (Extract, Transform, Load) process is among the most common (Kaoudi et al., 2018). ETL is still evidenced by have been in practice for relatively long time especially when it comes for establishing data warehousing systems. ETL starts with extracting information from multiple sources, and these sources could be pipelines, databases, flat files, API, and cloud services.

Figure 1 Top 8 Data Integration Techniques & Strategies (PixelPlex, 2020)

Subsequent to extraction, the data transforms; it can be scrubbed, standardized, and transformed into a format that can be accepted and processed by the target system. The information is moved into a single destination or the data centre where it can be stored and processed for insights sake where it can be either a Datawarehouse or data lake. ETL have always been discrete and deliberate activities that need to be planned and designed to ensure the right end product is achieved in terms of transformed data storage. Nevertheless, due to the increasing difficulties in analytics across platforms, organizations began to use more free methods of data integration as ELT, when data is loaded into the system and then transformed. This proves to be more efficient compared with the previous approach, and is more effective when used on extended use cases with massive volumes of text from various feed sources.

Another normal way of handling with the problems of data integration is by the help of APIs and Web services. APIs allow diverse systems to exchange information in real time, which makes them foundational to current approaches to data integration (Möller et al., 2017). APIs are the accepted pattern of sharing information between software applications with the intention of integrating different systems and platforms within an organization. By means of APIs it is possible to import information from different sources to analytics platforms without intricate conversions.

Because APIs come in a flexible structure, in other words they do not have a set type of integration, they become effective in spanning across platforms for analytics purposes for example cloud services, mobile apps and ERP systems. Furthermore, APIs play an essential role in automating data integrational while whereas limiting the role of a human being in the middle which in turn enhance the speed and accuracy of data integration. However, depending on an API also has its drawbacks, for instance where the question of data consistency arises as well as API versioning problems that arise when editing or interfacing to newer systems.

Data warehousing and data lakes comprise another huge area of the cross-platform integration. Data warehouses are productive facilities applied for the centralization of methodical and well-ordered information obtained from various sources that are used mainly for reporting. On the other hand, data lakes refer to the collecting of raw data, unorganized data, or semi-organized data in its natural format for analysis and processing of various types of data (Zhang et al., 2020). Data warehousing and data lakes are both used extensively for cross-platform, analytics because they enable organizations to place data from different sources in a single locale that is easily accessible.

Traditional data warehouses store highly defined and standardized kinds of data, such as transactional data, while an environment like a data lake is more

appropriate for big portions of semi-structured or non-structured information, including logs, snapshots, or data from sensors. The decision of opting for a data warehouse or a data lake depends on two factors; the nature of data for integration and the kind of analysis required in the organization. However, both approaches are now being used individually in an integrated hybrid system due to the realization of the strengths that will be accrued by each of the systems. Data warehouse can contain only structured data while a data lake contains unstructured data, and both are query able through the same interface for analysis.

The integration of real-time data is gradually becoming more of a necessity for organisations that are engaging in cross platform analysis. Faster and better decision-making is on the radar of many companies, thus real time integration and analysis of data is critical. Real-time data integration is the different from the batch processing method in which data is processed immediately when the data is produced (Agrawal et al., 2018). This way, technologies like stream processing and EDFA let the organization assess and process data and stream it from platforms such as social media, IoT, and transactions in real-time.

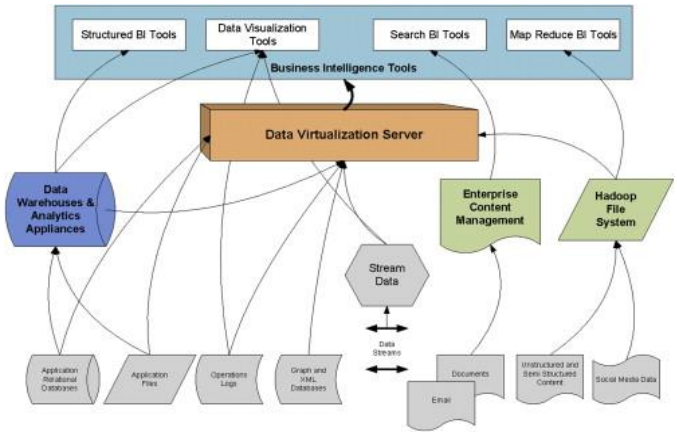

Figure 2 Data Integration Technique - an overview (ScienceDirect, 2020)

Techniques such as Apache Kafka and Apache Flink for stream processing allow real-time integration of data to give organizational entities a better shot at adapting to existing conditions and thus make better decisions in the long run. However true-time data integration also has some issues such as data synchronization across multiple platforms and managing large amount of data and complexity in data integration from different types of sources.

Another emerging type of data integration solutions is based on clouds, as many companies transfer their data and applications to the cloud. AWS, Azure, Google or IBM, for example – all these cloud platforms offer several tools and services aimed at the creation of integrated data flows across several platforms. Such services also come with connectors for well-known applications and databases, so to integrate data from different sources in an organization takes little effort.

Conveyance of integration also provides the feature of scalable and future flexibility so that it does not put limitations on the organizational capacity of handling mass quantity of data and also in expanding further its analytics portfolio (Camp et al., 2018). In the similar regard, cloud platforms are supportive of real time data processing and analysis, which makes them perfect for integration across platforms. However, cloud integration also poses loss of security as the organization is at risk of losing their data and also concerns of compliance since the data must meet the regulations of GDPR or HIPAA.

All in all, the methods to integrate data for cross-platform analytics are manifold and continue to emerge while organisations aggrandize demands to integrate and analyze data from different platforms. ETL and ELT, along with other techniques, are on demand, however, API-integration, real-time data processing, and cloud solutions are approaches that are coming to the forefront.

We can also conclude that every technique has their strengths and limitations and organization need to evaluate and decide what technique is best suited to their integration strategy. Therefore, new and more efficient cross-platform integration methods are expected to play the key role and meet the growing

user demand for cross-platform analytics to provide organizations with the greatest possible value from data.

Data Integration Techniques for Cross-Platform Analytics

Techniques of data integration for cross data analysis are a critical component in the ability of organizations in amalgamating large volumes of data from different systems for a coordinated analysis. As companies employ more of various platforms and technology tools, the leveraging of this data in real time and across these loops has emerged as a necessary component as the foundation for decision making and insights generation. Certainly, one of the methods used is ETL (Extract, Transform, Load). ETL facilitates the process of transferring data from various sources, preprocessing and structuring it to fit a required format as well as transferring it to a central system, the data warehouse, or the data lake.

This approach provides an easy way to approach the problem when working with structured data and for maintaining data integrity. For example, the extraction phase includes the extraction of information from one or more databases and/or one or more APIs or from other external platforms and the subsequent transformation phase, which deals with the cleaning and / or filtering and / or data normalization process (Chaturvedi et al., 2019). The last transfer of data is into a central data warehouse for processing to take place. An example of how to or form ETL process with data frames using Python's panda's library can be seen in the following code.

```
import pandas as pd

# Extract: Load data from two different sources
data1 = pd.read_csv('data_source1.csv')
data2 = pd.read_csv('data_source2.csv')

# Transform: Merge and clean the data
data_combined = pd.merge(data1, data2, on='common_column')  # Example of merging data
data_cleaned = data_combined.dropna()  # Remove rows with missing values

# Load: Save the cleaned, integrated data to a new CSV file
data_cleaned.to_csv('integrated_data.csv', index=False)
```

Another important approach applied for data integration is the use of APIs as the second crucial method after ETL. Real-time integration is thus made possible by APIs since systems can only inquire and exchange information as required. The most popular use of APIs is to extract information from outside platforms like cloud solutions or other applications to be used in an analytics environment. In this method, the data update is equally constant, therefore useful in real-time analysis and decision-making. Integrating ETL with APIs can greatly enrich an organization's capability to amalgamate data across systems to a central platform for real-time analytics.

Case Studies or Examples

It is found that in the field of cross-platform analytics several companies have integrated their data to improve the operation and decision making. A good example is Netflix that depends on data integration as one of the key strategies in offering customers' customized services and content delivery (Mangaroska et al., 2019). The system implemented by Netflix for data integration include ETL processes, APIs, and real-time data integration processes to gather and process large amount of data from several diverse resources.

A key piece of the company is its recommendation engine, which is largely responsible for the interactions of its users; the data from these interactions through multiple contexts such as devices, streaming platforms, extended data from social media and content rating engines are utilized in the recommendation system. The sources of data of Netflix are both internal data warehouse and of the sophistically AWS services. Through APIs, Netflix is then able to, with real time data processing from the mobile applications, smart TVs, and other connected devices. Consequently, the company can make specific content suggestions to the users of the application, thus enhancing the user loyalty and usage rates.

In the same way, through a complex data integration system, Amazon can effectively process and analyse data from the e-Commerce site, customers, products,

stocks, logistics and even other third parties (Zelenkauskaite, 2017). Data integration in Amazon: It has a hybrid mode of integration, where they bring data from within their company, AWS services, on-premises as well as from external APIs. AWS enables this integration through offerings such as Amazon Redshift, AWS Glue and Kinesis, through which Amazon can integrate the collecting, transforming and real time processing of the data.

For example, through real-time data integration, Amazon can immediately identify its customers' purchases, determine its stock status and adjust its product's prices according to its demand. Amazon has made sure that it presents relevant data to give customers a variety of the right products at the right price with satisfactory operations (Majchrzak et al., 2017). The factor that has placed the company in a very strategic position is the capability effectively to assimilate and process large volumes of information within a short span of time.

Uber is another fabulous example that execute the idea of data integration and it is applied in efficient formulation of car service, vigilance over the riders and in enhancing the satisfactory level of the clients. Uber records information from its mobile application, Global positioning system, payment interfaces, client feedback amongst others. The firm incorporates this datum through cloud services and APIs to generate real-time ride fare, passenger and drivers' matching, and optimization of the nearest driver. Another insight, which Uber can produce with the help of information from diverse sources, is demand in specific regions and prices will be changed with reference to it, which gives drivers a motivation to satisfy customers' needs in the shortest time (Oliveira et al., 2017). Another aspect, which has always proved useful in Uber's case, is the capability to integrate real-time data to extend operations internationally and offer a consistent service across different applications.

Netflix, Amazon, and Uber are great examples in which various techniques of data integration such as ETL process, API, real time data processing etc., help an organization to improve the possibility of decision making, creating customer value proposition, and operational efficiency.

Through the accumulation of data from different channels, these firms are also able to further enhance their organizational effectiveness and, more importantly, be able to serve their customers better and, as such, consider themselves as industry pioneers. All of these companies have effectively implemented high-functioning data integration and management in quite a competitive economy. These firms mean that with structured and unstructured data sourcing and ingestion, they harness valuable insights that cause innovation and business value.

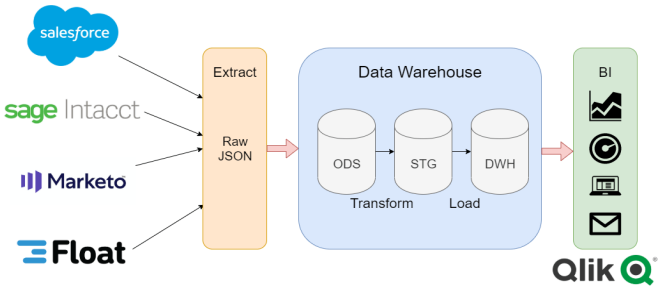

Figure 3 Data Integration Tools & Techniques (Analytics8, 2020)

III. Results and Discussion

Cross-platform data integration techniques are usually very effective when it comes to organisational improvement because they bring better efficiency, improved decisions, and better customers. Companies can access and curate information from a variety of sources and store them all in one platform where they can provide additional capabilities for analysis. For example, the Netflix service, which uses data from different platforms, leads to unique recommendations because of increased user retention. Arguably, the same way Amazon harnesses e-commerce data and links it with information about inventory and logistics guarantees some products' availability when customers demand them and vice versa enhances sales of some products and efficiency

in the company's operations (Shah et al., 2019). These examples basically speak for themselves and denote the vast potential of the data integration strategies highlighted here.

Also, a real-time integration of data, which has been observed for such corporations like Uber, provides the opportunity to make instant decisions at a given moment. The fact that Uber has the power to change the price of its services frequently and in real time is made possible by the feed from its App, GPS and payment interfaces. This integration is the best way for Uber to ensure that the rate of supply of drivers and the rate of demand of passengers' service is seamless. Real-time data processing provides competitive advantage to any business since organisational response to dynamic environment is much faster.

Nevertheless, like in any other technologies, cross-platform data integration brings benefits for companies together with definite difficulties. One of the primary challenges of the process is the handling of big data which could have been derived from various sources, formats included. Also, viability in data privacy and security is paramount in the integration process since including delicate data from multiple systems attracts vulnerability to scams (Biswas et al., 2020). However, there is an importance associated with data integration, as it seems to be one of the key drivers of the long-term success when properly integrated, despite numerous challenges it may pose.

IV. Conclusion

Data integration techniques are very useful in cross platform analysis and is a strong weapon in the hands of any business organization to integrate large chunk of data for the purpose of better analysis and decision making. Such working principles as machine learning, big data, and deep learning are illustrated by several examples, such as Netflix, Amazon, and Uber. Despite the challenges mentioned above such as different data format and security the potential benefits are substantial making data integration a strategic necessity for organizations which aim to remain innovative in an age of growing data relevance. Amidst various layers of cross-platform analytics, proper integration of the data forms the core strategy necessary for success.

V. REFERENCES

[1]. Agrawal, D., Chawla, S., Contreras-Rojas, B., Elmagarmid, A., Idris, Y., Kaoudi, Z., ... & Troudi, A. (2018). Rheem: enabling cross-platform data processing: may the big data be with you!. Proceedings of the VLDB Endowment, 11(11), 1414-1427. https://doi.org/10.14778/3236187.3236195

[2]. Banović, V., & Sorić, I. (2019, May). General solution framework for management and monitoring of cross-platform data processes. In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 1288-1292). IEEE. https://doi.org/10.23919/MIPRO.2019.8757065

[3]. Biswas, S., Sharma, L. K., Ranjan, R., & Banerjee, J. S. (2020). Go-COVID: an interactive cross-platform based dashboard for real-time tracking of COVID-19 using data analytics. J Mechanics Continua Math Sci, 15, 1-15. https://d1wqtxts1xzle7.cloudfront.net/63740269/COVID-19_JMCMS_JSB_June_202020200625-3976-1g3mm8u-libre.pdf?1593159465=&response-content-disposition=inline%3B+filename%3DGO_COVID_An_Interactive_Cross_Platform_B.pdf&Expires=1731224827&Signature=SV2gC7C5K9jkGwoqVcxnHM86y-j-EBaPsbvLk2Aj8gg4CEa3d7lsmyxrncYW123t4a

D0KGag-mI90vXkAmjl~7rDvMJ4II1d82ltOXAC56BQJqi7HCjzOdPqcHzcvAvN3RNvRBEslwZ503RQ0SYFNTFNiyqJ1ZrEVmKi-jCJC3jci1ff0j~HNJznwWfobSL0hVTJIkfo8XfAkDbaYNTctc5L7O0zzcv-8l2Msf3s08PyYyJOdRBt6wGAFg73I32azUkxxLoBXZcniIiBb0oRoxFH71RxUSNQCYQhYAWCPPZRg603jJpXY18lH4jhALYRjF-y1YDYCBlSaIq~L5OWew__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[4]. Camp, B., Mandivarapu, J. K., Ramamurthy, N., Wingo, J., Bourgeois, A. G., Cao, X., & Sunderraman, R. (2018). A new cross-platform architecture for epi-info software suite. BMC bioinformatics, 19, 1-8. https://doi.org/10.1186/s12859-018-2334-8

[5]. Chaturvedi, K., & Kolbe, T. H. (2019). Towards establishing cross-platform interoperability for sensors in smart cities. Sensors, 19(3), 562. https://doi.org/10.3390/s19030562

[6]. Kaoudi, Z., & Quiané-Ruiz, J. A. (2018, April). Cross-platform data processing: use cases and challenges. In 2018 IEEE 34th International Conference on Data Engineering (ICDE) (pp. 1723-1726). IEEE. https://doi.org/10.1109/ICDE.2018.00223

[7]. Majchrzak, T., & Grønli, T. M. (2017). Comprehensive analysis of innovative cross-platform app development frameworks. http://hdl.handle.net/10125/41909

[8]. Mangaroska, K., Vesin, B., & Giannakos, M. (2019, March). Cross-platform analytics: A step towards personalization and adaptation in education. In Proceedings of the 9th international conference on learning analytics & knowledge (pp. 71-75). https://doi.org/10.1145/3303772.3303825

[9]. Möller, S., Prescott, S. W., Wirzenius, L., Reinholdtsen, P., Chapman, B., Prins, P., ... & Crusoe, M. R. (2017). Robust cross-platform workflows: how technical and scientific communities collaborate to develop, test and share best practices for data analysis. Data Science and Engineering, 2(3), 232-244. https://doi.org/10.1007/s41019-017-0050-4

[10]. Oliveira, F., Suneja, S., Nadgowda, S., Nagpurkar, P., & Isci, C. (2017, December). Opvis: extensible, cross-platform operational visibility and analytics for cloud. In Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference: Industrial Track (pp. 43-49). https://doi.org/10.1145/3154448.3154455

[11]. Shah, K., Sinha, H., & Mishra, P. (2019, March). Analysis of cross-platform mobile app development tools. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-7). IEEE. https://doi.org/10.1109/I2CT45611.2019.9033872

[12]. Zelenkauskaite, A. (2017). Remediation, convergence, and big data: Conceptual limits of cross-platform social media. Convergence, 23(5), 512-527. https://doi.org/10.1177/1354856516631519

[13]. Zhang, S., Shao, J., Yu, D., Qiu, X., & Zhang, J. (2020). MatchMixeR: a cross-platform normalization method for gene expression data integration. Bioinformatics, 36(8), 2486-2491. https://doi.org/10.1093/bioinformatics/btz974