International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Finding the Right Data Analytics Platform for Your Enterprise

Laxminarayana Korada[1], Vijay Kartik Sikha[2], Satyaveda Somepalli[3]

[1]Email: *laxminarayana.k[at]gmail.com*
ORCID: 0009-0001-6518-0060

[2]Email: *vksikha[at]gmail.com*
ORCID: 0009-0002-2261-5551

[3]Email: *satyaveda.somepalli[at]gmail.com*
ORCID: 0009-0003-1608-0527

Abstract: ***While data is an asset for businesses, the rapid increase in data generation presents challenges in storage, management, and processing. The proliferation of IoT devices, modernization of legacy applications, a global user base of 4.9 billion social media users (Wong, 2023), and the rise of 5G mobile applications generate vast amounts of unstructured or semi-structured data. While reliable and cost-effective storage solutions exist, enterprises rely on dedicated teams led by a Chief Data Officer (CDO) to manage and optimize data. Centralized data models give these teams authority over data management, while decentralized models allow for organization-level strategies with delegated management. Developing in-house solutions for data organization and analysis, which can reach petabytes or exabytes, is challenging. Numerous platforms, such as Microsoft Fabric, AWS Analytics, Oracle Data Analytics, and Google Analytics, streamline transforming raw data into actionable insights. CDO is crucial in establishing data governance and management policies that support AI adoption, ensure security, and align with open standards. Data analytics guides businesses towards their vision, and choosing the right platform enhances its impact.***

Keywords: Data, data centralization, decentralized data, analytic platforms, role of chief data officer

1. Introduction

Data is a valuable commodity in today's business landscape, and enterprises are well-positioned to capitalize on this asset. A report projects that by 2025, more than 75 billion IoT devices will be connected and utilized (Statista, 2022). However, when gathering significant amounts of data from various sources, including cloud repositories and on-premise data centers, the volume of data stored becomes staggering.

The true challenge is translating this raw data into meaningful insights that drive informed decision-making and accelerate development. Only 26.5% of organizations (Business Wire, 2022) consider themselves excellent at this. The need to use an appropriate data analytics platform has become critical as businesses struggle with the intricacies of data management and analysis.

Businesses face a maze of alternatives, such as hybrid architectures, cloud-native platforms, and traditional on-premises solutions, each with its own advantages and disadvantages. The objective here is to simplify the data analytics landscape, enabling organizations to select the platform that best aligns with their unique requirements, infrastructure, and long-term goals.

2. The Data Landscape of an Enterprise

The modern enterprise data landscape is a multifaceted ecosystem that encompasses extensive collection of data sources and is continually expanding. The healthcare data alone is projected to reach 180 zettabytes by 2026 (Kumar, 2024). To illustrate the vastness of this amount, one may consider comparing it to the storage capacity of high-density hard drives. In 2024, the average hard drive capacity is around 20 terabytes and storing 160 zettabytes would require 8 billion hard drives. The generation of voluminous data can be attributed to data generated from electronic health records, imaging, genomic data, wearables, pharmacies, patient portals, and insurance systems. The data sources are divided into two broad categories.

a) Structured data sources that include electronic health records, clinical documentation, medical imaging systems, laboratory information systems, and health insurance claim data.
b) Unstructured data sources which include clinical notes, medical images, telemedicine consultation, patient data generated from wearable devices, public health surveys, and text-based reports.

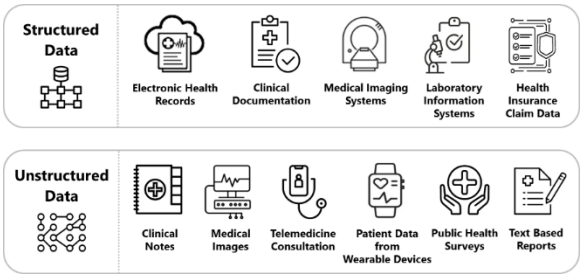

Figure 1: Healthcare Data Sources

While this forecast pertains to the healthcare industry, the amount of data generated and stored across all industries is inestimable. This data necessitates further processing and transformation into an appropriate format to run analytics, generate reports, and derive insights that can improve treatment outcomes.

According to McKinsey, Global Institute, manufacturing is one of the most prolific data-generating industries, generating an average of 1.9 petabytes of data every year. Similarly, the

global banking industry generates around 2.5 quintillion bytes of data every day (ScikIQ, 2023).

Identifying the most suitable storage option for this data represents the primary challenge. A multitude of storage options, each tailored to address distinct requirements, store a variety of data sets. Although still significant, traditional relational databases, renowned for their structured data storage and robust querying capabilities, are gradually being complemented by other solutions.

In light of the growing prevalence of semi-structured and unstructured data, NoSQL databases have gained significant prominence as an effective means of managing various types of data formats. These are often complemented by object-based blob storage systems, which can hold large binary data items. Despite their utility, storing and managing such data presents a formidable challenge that is often underestimated.

The following are the common issues that emerging enterprises typically encounter as the volume of data grows substantially (Chesney, 2023):

a) Data silos among departments such as operations, HR, marketing, and business, seem harmless. However, siloed data can create a barrier that can affect an organization's decision-making ability. Siloed, duplicate, or incomplete data sources can lead to a lack of trust when the data is funneled into enterprise systems.
b) Data is often ingested by customers or from other sources; however, it is transferred and shared with many internal groups. This leads to the creation of many data owners who view the same data differently. An improper enterprise data management program can lead to process inefficiencies, inadequate data access, or security controls, and the data can lose its integrity.
c) If the data foundation is unreliable, the data processed is prone to errors, ultimately leading to unreliable outputs. In addition, the process of revalidating and correcting the data will consume valuable time and resources, ultimately resulting in unproductive outcomes.
d) Legacy systems and antiquated infrastructure present twofold challenges: they cannot be expanded as needed and incur significant operational expenses.
e) The absence of standardized data formats, naming conventions, and data management processes can lead to integration challenges and negative downstream impacts.

The challenge for businesses lies in effectively coordinating a multitude of data sources, formats, and storage systems within a complex framework. By integrating these disparate elements, companies can unlock the value of their data assets and derive valuable insights that contribute to informed decision-making and business growth.

3. Centralization vs. Distributed Data

Data, often referred to as the "new oil, " generates annual revenues exceeding $1 trillion. The existing data storage, management, and utilization infrastructure of numerous enterprises, however, can result in data breaches and cybercrimes.

Centralized and distributed database management are two approaches that are critical for enterprises dealing with data analytics. These approaches can significantly influence an enterprise's ability to extract value efficiently from its data assets.

The decision to adopt a distributed architecture across multiple platforms or to centralize data within a single repository presents a complex challenge with various pros and cons. Following section covers the intricacies of this architectural conundrum.

3.1 Data Centralization: The Unified Approach

Data centralization is a unified and consolidated method for administering an enterprise's data assets. Centralizing all data in a single, centralized repository enables businesses to streamline management, expedite access to data, and facilitate efficient querying and analysis. This strategy can lead to significant cost savings and resource optimization.

The centralization of data poses distinct challenges. Developing and maintaining a dependable, centralized data infrastructure is an extensive undertaking that typically requires substantial investment in hardware, software, and personnel. Businesses must guarantee that their central repository is scalable, accessible, and current with the relentless influx and rapid pace of ingestion. The rigidity of the data store can ultimately impede digital transformation and the progress of contemporary application development, thereby jeopardizing the objectives of the organization.

3.2 Distributed Data: The Decentralized Paradigm

Data decentralization across multiple platforms and locations is the hallmark of a distributed data architecture as opposed to the centralized approach. This paradigm gives businesses the required flexibility and agility to employ specialized data storage solutions designed for specific use cases or workloads. For instance, transactional data may reside in a relational database, whereas unstructured data, such as media files or log entries, can be stored in object-based repositories.

Despite the benefits that this approach offers, it also presents certain complexities. One of the main challenges is the distribution of data, which can make the search and analysis process more complicated. For enterprises that want to ensure easy access and analysis of data that is dispersed across multiple platforms, it is essential to have strong data integration and visualization plans in place. Implementing advanced data pipelines, orchestration tools, and federated querying techniques can further complicate the architecture, but they are often necessary to achieve the desired results.

4. The Role of Analytics Platforms

In the digital age, analytics platforms have become pivotal in transforming raw data into actionable insights. These platforms enable businesses to make data-driven decisions, optimize operations, and enhance customer experience. The public cloud analytics platforms that have significantly influenced the industry are Microsoft Fabric, AWS Analytics, Google Analytics, and Oracle Analytics Cloud.

4.1 Microsoft Fabric

Microsoft Fabric is an end-to-end data analytics platform designed to provide a comprehensive suite of services for data integration, warehousing, and visualization. It seamlessly integrates with other Microsoft services within Azure, Power Platform, and Dynamics 365.

4.1.1 Features and Functionalities

a) **Data Integration and ETL (Extract, Transform, Load):** Microsoft Fabric offers robust data integration capabilities, allowing users to connect to various data sources, both on-premises and in the cloud. It supports ETL processes to ensure data is cleansed, transformed, and ready for analysis (Microsoft, 2021).
b) **Data Warehousing**: The platform provides scalable data warehousing solutions through Azure Synapse Analytics. This allows businesses to store and manage large volumes of data efficiently (Vossen, 2021).
c) **Advanced Analytics and AI**: With built-in AI and machine learning tools, Microsoft Fabric enables predictive analytics, helping organizations forecast trends and make proactive decisions (Henschen, 2020).
d) **Visualization and Reporting**: Integrated with Power BI, Microsoft Fabric offers extensive data visualization capabilities, allowing users to create interactive dashboards and reports that enhance data comprehension and decision-making (Gartner, 2022).

4.1.2 Impact on Businesses

Microsoft Fabric presents a comprehensive solution for businesses that require a unified platform for their data management needs. Through its integration with other Microsoft services, it enhances productivity and collaboration within organizations. Companies can utilize robust analytics capabilities to gain deeper insights and optimize their operations, ultimately improving customer engagement (Rajgopal & Mooney, 2021).

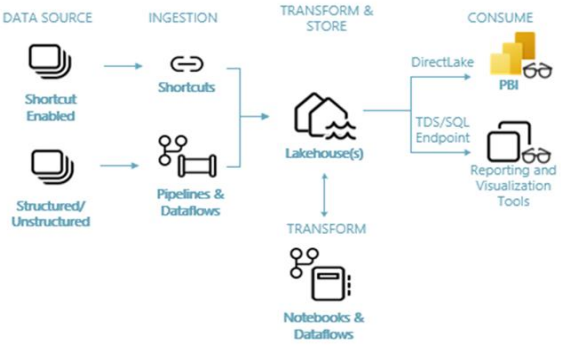

Figure 2: Azure Fabric Data Pipeline Diagram

The image illustrates a data pipeline diagram, detailing the flow and transformation of data from its source to its consumption utilizing Microsoft Fabric architecture. Here's a breakdown of the content:

1) Data Source
2) Ingestion
3) Transform & Store
4) Consume

4.2 AWS Analytics

Amazon Web Services (AWS) offers a comprehensive suite of analytics services that enable businesses to process and analyze data on a large scale. This suite includes Amazon Redshift, AWS Glue, Amazon Quick Sight, and AWS Lake Formation, among other services.

4.2.1 Features and Functionalities

1) Data Warehousing: Amazon Redshift provides a fast, scalable data warehousing solution that enables complex queries and analysis across large datasets. It integrates with various AWS services, enhancing its functionality and flexibility (Madden, 2020).
2) Data Integration and ETL: AWS Glue is a fully managed ETL service that makes it easy to prepare and load data for analytics. It automates the process of discovering, cataloguing, and transforming data from diverse sources (Amazon, 2021).
3) Visualization and Reporting: Amazon Quick Sight allows users to create interactive dashboards and visualizations. It supports natural language queries, making it accessible to non-technical users (AWS, 2021).
4) Data Lakes: AWS Lake Formation simplifies the process of setting up and managing data lakes. It allows businesses to store, catalog, and secure large volumes of data from various sources (Jassy, 2019).

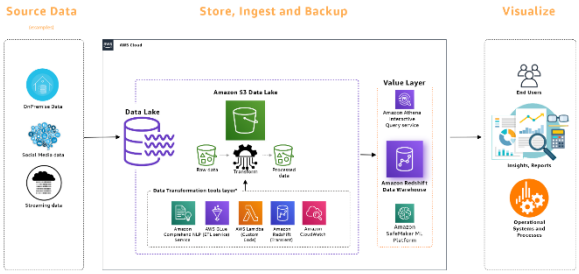

Figure 3: Data Processing & Analytics Architecture using AWS

The image illustrates comprehensive data processing and analytics architecture using AWS services. Here's a detailed breakdown of its content:

- Source Data
- AWS Cloud Components
- Value Layer
- Visualize

4.2.2 Impact on Businesses

AWS Analytics provides businesses with a robust and scalable platform for managing and analyzing data. Its wide range of services caters to different analytical needs, from data warehousing to real-time analytics. By leveraging AWS Analytics, businesses can gain deeper insights, improve operational efficiency, and drive innovation (Leach, 2020).

4.3 Google Data and Analytics

Google provides extensive options for data storage, processing, analytics, and visualization. The unified data platform of Google Cloud allows organizations to establish a robust AI-ready data foundation with BigQuery, while several other services facilitate data lake modernization.

4.3.1 Features and Functionalities

1) **Data Warehouse:** Google BigQuery is a highly efficient data analysis tool designed to process extensive datasets with remarkable speed. It enables users to derive actionable insights from raw data using SQL queries. This powerful platform serves as an indispensable resource for sophisticated data analytics and business intelligence. Google Cloud Platform (GCP) offers support for data migration from various services, such as Terradata, Oracle, Snowflake, and Apache Hive.
2) **Data processing:** Google Dataproc, the principal offering for data processing, is a robust cloud-based service designed for processing large-scale data using Apache Hadoop and Spark. It facilitates efficient management and analysis of big data by automating cluster creation, scaling, and management. This platform is essential for organizations seeking to streamline their data processing workflows and enhance their analytical capabilities.
3) **Integration with Google Services**: The Google Cloud Platform integrates seamlessly with other Google services, such as Google Ads and Google Search Console, providing a holistic view of digital marketing efforts (Ledford, 2021).
4) **Traffic Analysis**: Google Analytics provides detailed reports on website traffic, including metrics such as page views, sessions, bounce rates, and user demographics. This helps businesses understand their audience and how users interact with their site (Clifton, 2012).

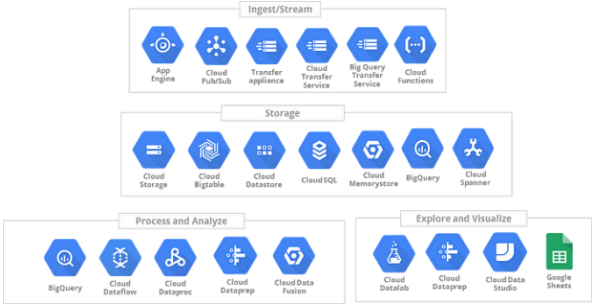

Figure 4: Google Data and Data Analytics Products

The diagram depicted above illustrates the comprehensive suite of services offered by Google to meet the extensive data management requirements of a business (Yagci, 2022). These services encompass every aspect of the data pipeline, and they are designed to seamlessly integrate with other complementary solutions. The following are some of the key services:

- BigQuery
- Dataflow
- Pub Sub
- DataProc

4.3.2 Impact on Businesses

Google Cloud's Data Analytics services offer businesses a serverless, fully managed platform for generating insights at any scale. The integration of these services with Vertex AI allows businesses to achieve their AI objectives. Google Analytics is essential for businesses looking to optimize their online presence. By providing detailed insights into user behavior and website performance, it enables companies to make data-driven decisions that enhance user experience and increase conversions. Its integration with other Google services allows comprehensive digital marketing analysis, leading to more effective strategies (Batra, 2021).

4.4 Oracle Analytics

Oracle Analytics is a cloud-native service which can be set up either on-premise or in the cloud. Oracle Analytics Cloud (OAC) is a comprehensive analytics platform that harnesses advanced efficiencies of cloud technology and provides a suite of services for data preparation, visualization, reporting, and advanced analytics. It is designed to cater to the needs of both technical and non-technical users, offering self-service analytics capabilities.

4.4.1 Features and Functionalities

1) **Data Preparation**: OAC simplifies data preparation with automated data profiling, enrichment, and transformation. This enables users to prepare data for analysis without extensive technical expertise (Oracle, 2020).
2) **Visualization and Reporting**: OAC offers a wide range of visualization tools, including interactive dashboards, graphs, and charts. Users can create compelling reports that facilitate data-driven decision-making (Gartner, 2021).
3) **Machine Learning and AI**: The platform integrates machine learning algorithms that enable predictive analytics and anomaly detection. These capabilities help organizations uncover hidden patterns and insights in their data (Nucleus Research, 2021).
4) **Collaboration and Sharing**: OAC promotes collaboration by allowing users to share insights and reports easily. It supports integration with other Oracle applications, enhancing data accessibility and usability (Oracle, 2021).

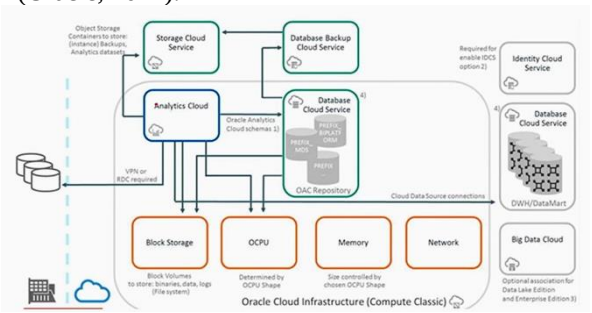

Figure 5: Oracle Analytics Cloud Architecture

This architecture diagram illustrates the integration of the Oracle Analytics Cloud (OAC) within Oracle Cloud Infrastructure (OCI) and its connection to external database services, such as DataLake/DataMart, as well as other edge applications and services like BigData ("Getting Started With Oracle Analytics Cloud", 2019). The following is a list of the various solution components depicted in the diagram:

- Storage Cloud Service: Oracle's Object Storage is a platform designed to store or retrieve data from the internet or within the cloud. It can be accessed by Oracle Cloud Infrastructure resources in a VCN (Virtual Cloud Network) through a service gateway.
- Database Backup Cloud Service: Oracle's offering of a scalable, secure, on-demand storage solution for backing up Oracle databases to Oracle Cloud. It serves as an

additional backup strategy by storing the data in public cloud off-premises storage location.

- Database Cloud Service: Oracle Base Database Service provides Oracle Databases hosted on Oracle Cloud Infrastructure. It offers single-node DB systems on either bare metal or virtual machines and 2-node RAC DB systems on virtual machines. (Oracle)
- Block Storage: Oracle's Block Storage boasts its design centered around durability and security of data. It is an iSCSI Ethernet protocol-based data storage type leveraged to deliver performance and functionality capabilities and performance like on-prem SANs (Storage Area Networks). Block volumes can directly be created and embedded and attached to OCI compute instances.

OCI user access and security are handled via Oracle's Identity Cloud Service for the best results.

4.4.2 Impact on Businesses

Oracle Analytics Cloud enhances business agility by providing intuitive and powerful analytics tools. Its machine learning capabilities allow businesses to anticipate market changes and customer behavior, leading to more informed strategic decisions. The platform's self-service nature democratizes data access, empowering employees at all levels to contribute to data-driven initiatives (Tschakert et al., 2019).

4.5 Comparative Analysis

While all four platforms offer robust analytics capabilities, they serve different purposes and target audiences. Microsoft Fabric is ideal for organizations seeking an integrated data analytics solution that combines data warehousing, AI, and visualization. Oracle Analytics Cloud is suited for businesses that need advanced analytics with strong machine learning capabilities and self-service features. Google Analytics is tailored for web analytics, providing insights into website traffic and user behavior while GCP Analytics services provide data querying capabilities at any scale. AWS Analytics, with its comprehensive suite of services, is perfect for businesses looking for a scalable and flexible analytics platform that can handle diverse analytical needs.

5. Role of Chief Data Officer

The Chief Data Officer (CDO) plays a crucial role as organizations increasingly rely on data for strategic decision-making. One of the CDO's primary responsibilities is to guide their teams in utilizing and managing large amounts of data with suitable data storage and analytics solutions. As the guardian of the organization's data assets, the CDO is pivotal in choosing the right analytics platform that can deliver high-performance analytics and manage the company's data volumes for both current and future needs. The chosen platform must ensure seamless data integration from various sources while adhering to the company's data governance guidelines and regulatory compliance requirements. Additionally, the CDO must evaluate whether the platform can accommodate the organization's sophisticated analytical requirements, such as machine learning and predictive modeling.

5.1 Factors a CDO Should Consider When Choosing the Right Data Analytics Platform

When determining the optimal platform for managing and storing volumetric data, a CDO should consider the following factors to ensure efficient data management.:

1) **Understanding the Business Requirements** An analytics solution must align with the organization's existing and future business needs. The CDO should conduct thorough analysis and research to comprehend the business's core objectives and list the desired outcomes. These objectives should then be broken down into measurable analytic goals. The chosen analytics platform should offer reporting capabilities and features that facilitate the achievement of these business goals.
2) **Pricing** Making a well-informed decision requires understanding all costs, including hidden fees, subscriptions, and potential price increases associated with the solution. Different analytics platforms have varying pricing structures, and a thorough understanding is essential to avoid future financial challenges.
3) **User Interface and Visualization** A user-friendly interface is crucial for maximizing user adoption and driving better outcomes from technology investments. Self-service analytics should cater to a wide range of user personas, including those with non-technical backgrounds. Strong data visualization capabilities are also important, as poor visuals can negatively impact user adoption and the overall effectiveness of the platform.
4) **Advanced Analytics** Advanced analytics, which includes recognizing patterns in data and predicting future trends, events, and outcomes, is essential for making data-driven decisions. This capability allows organizations to develop advanced statistical models that can future-proof their business by providing contextualized insights.
5) **Integration** It is essential to determine whether the organization requires a standalone solution or an integrated solution that can handle data from multiple sources. While there are many standalone solutions, integrated solutions often offer more comprehensive benefits. The chosen system should seamlessly integrate with existing infrastructure, applications, and external data sources. Additionally, the complexity of data migration should be considered.
6) **Multiple Sources of Data** The analytics platform should be able to combine and analyze semi-structured, structured, and unstructured data from multiple sources. This capability enables the platform to provide a simplified version of the data on a single dashboard, offering 360-degree visibility of various business functions.
7) **Security** Ensuring advanced security standards to safeguard data assets is critical when selecting a data analytics platform. The CDO must ensure that the platform complies with security protocols and can protect sensitive data. Additionally, understanding the implications of providing data access to users outside the organization's network perimeter is essential.

By carefully considering these factors, a CDO can select a data analytics platform that not only meets the organization's

current needs but also supports its future growth and strategic objectives.

6. Conclusion

Implementing an analytics platform is not solely a technical necessity but also a strategic imperative that is essential for an organization's success in the current data-driven environment. It is important for organizations to address this imperative comprehensively by thoroughly assessing their data environment, analytics requirements, scalability demands, security measures, and user experience objectives. Innovation and competitive advantage depend on Chief Data Officers being empowered to lead these important initiatives.

Having the right analytics platform can transform an organization's underutilized data assets into valuable insights that drive customer-focused strategies, operational efficiency, and business growth. In a data-centric market, companies that prioritize this and align their platform choices with their overall business goals will be well-positioned to outperform their competitors.

References

[1] "IoT Devices Installed Base Worldwide 2015-2025 | Statista. " Statista, 27 July 2022, www.statista. com/statistics/471264/iot-number-of-connected-devices-worldwide.
[2] "NewVantage Partners Releases 2022 Data And AI Executive Survey. " 2022. Business Wire. https: //www.businesswire. com/news/home/20220103005036/en/NewVantage-Partners-Releases-2022-Data-And-AI-Executive-Survey.
[3] Kumar, Naresh. n. d. "101 Guide to Healthcare Data Integration for Enterprises. " Zuci Systems. Accessed July 12, 2024. https: //www.zucisystems. com/blog/healthcare-data-integration/.
[4] "Banking Analytics: Top 10 Use Cases of Data Analytics in Banking. " 2023. ScikIQ. https: //scikiq. com/blog/top-10-use-cases-of-data-analytics-in-banking/.
[5] Chesney, Scott.2023. "5 Enterprise Data Management Challenges and How to Overcome Them. " RCG Global Services. https: //rcgglobalservices. com/blog/5-enterprise-data-management-challenges-and-how-to-overcome-them.
[6] Amazon. (2021). AWS Glue: Overview. Amazon Web Services.
[7] AWS. (2021). Amazon QuickSight: Interactive Dashboards. Amazon Web Services.
[8] Batra, R. (2021). The Impact of Google Analytics on Digital Marketing Strategies. Journal of Digital Marketing, 12 (3), 45-59.
[9] Clifton, B. (2020). Advanced Web Metrics with Google Analytics. John Wiley & Sons.
[10] Gartner. (2021). Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner Inc.
[11] Gartner. (2022). Microsoft Power BI for Analytics and Business Intelligence. Gartner Inc.
[12] Henschen, D. (2020). Microsoft Fabric's Role in Advanced Analytics. Enterprise Strategy Group, 9 (2), 23-34.
[13] Jassy, A. (2019). AWS Lake Formation: Simplifying Data Lakes. Amazon Web Services.
[14] Ledford, J. (2021). Google Analytics: Expert Techniques and Practices. O'Reilly Media.
[15] Leach, P. (2020). The Scalability of AWS Analytics. Journal of Cloud Computing, 7 (1), 34-49.
[16] Madden, S. (2020). High-Performance Data Warehousing with Amazon Redshift. Journal of Data Engineering, 14 (3), 67-81.
[17] Microsoft. (2021). Microsoft Fabric: Integrating Data and AI. Microsoft Corporation.
[18] Nucleus Research. (2021). The Benefits of Oracle Analytics Cloud for Predictive Analytics. Nucleus Research Inc.
[19] Oracle. (2020). Oracle Analytics Cloud Data Preparation Capabilities. Oracle Corporation.
[20] Oracle. (2021). Enhancing Collaboration with Oracle Analytics Cloud. Oracle Corporation.
[21] OAC—Getting Started With Oracle Analytics Cloud. xmlandmore. blogspot. com/2019/02/oacgetting-started-with-oracle. html.
[22] July2020. "Infrastructure and Platform Services (IaaS/PaaS) Billing Guide. " Oracle Help Center, docs. oracle. com/en/cloud/get-started/subscriptions-cloud/ocpib/key-billing-terminology. html. Rajgopal, S., & Mooney, J. (2021). Leveraging Microsoft Fabric for Data-Driven Decision Making. International Journal of Business Intelligence Research, 12 (1), 56-72
[23] Tschakert, N., Kokina, J., Kozlowski, S., & Vasarhelyi, M. (2019). The next frontier in data analytics. Journal of Accountancy, 227 (4), 28-35.
[24] Vossen, G. (2021). Scalable Data Warehousing with Azure Synapse Analytics. Journal of Data Science and Analytics, 15 (2), 112-130.
[25] Wong, Belle, JD. "Top Social Media Statistics and Trends of 2024. " Forbes Advisor, 18 May 2023, www.forbes. com/advisor/business/social-media-statistics.
[26] Yagci, Erim. "Modern Data Warehouse on Google Cloud Platform-Erim YAGCI-Medium. " Medium, 4 Jan.2022, medium. com/[at]erimyagci/modern-data-warehouse-on-google-cloud-platform-d91cc0b3b7c.