iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.57219

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Cloud-based Data Analytics for Business Intelligence

Shalin Dashora[1], Navneet Gupta[2], Milind Pal Singh Tanwar[3], Abhinav Singh[4], Sukhmeet Kaur[5]
Apex Institute of TechnologyChandigarh University Punjab, India

Abstract: ***In the era of data-driven decision-making, organizations are turning to cloud-based data analytics for business intelligence to overcome the limitations of traditional on-premises systems. This paradigm shift offers the promise of scalable, agile, and advanced analytics capabilities.***
This paper explores the landscape of cloud- based data analytics for business intelligence by investigating existing systems, challenges, and opportunities. The study first examines leading cloud platforms such as Amazon Web Services (AWS) Redshift, Microsoft Azure Synapse Analytics, Google BigQuery, and Snowflake[11]. It evaluates their features, scalability, and integration options to assess their suitability for modern BI needs.
Moreover, the research identifies critical challenges in transitioning to cloud-based analytics, including data integration complexities, security concerns, and cost management. The integration of advanced analytics techniques, such as machine learning and AI, into cloud- based environments is also explored. The study delves into the benefits and challenges of predictive analytics, anomaly detection, and other emerging capabilities that empower organizations to extract deeper insights from data. Furthermore, hybrid cloud architectures, which combine on-premises infrastructure with cloud resources, are investigated. Strategies for seamless data integration and workload distribution are discussed, enabling organizations to strike a balance between performance and data governance.
Keywords: Cloud-based data analytics, business intelligence, cloud platforms, data visualization, scalability, security, AI, hybrid clouds.

I. INTRODUCTION

The traditional landscape of business intelligence (BI) anddata analytics is undergoing a significant transformation with the advent of cloud computing. This transformation stems fromthe challenges inherent in conventional on-premises BI systems, which struggle to meet the demands of modern data-driven organizations.

The problem at hand revolves around the limitations posed by these traditional systems and the imperative to transition towards cloud-based data analytics foreffective business intelligence.

The primary challenge lies in the inability of on-premises BI systems to handle the burgeoning volumes of data generated by organizations today. As data complexity and diversity increase, traditional systems exhibit shortcomings in terms of scalability, adaptability, and accessibility.

The lack of seamless scalability often results in performance bottlenecks and sluggish query responses, hindering the timely extraction[3] of insights. Moreover, these systems are ill-equipped to cater to the dynamic nature of business requirements, making it difficult to swiftly integrate new data sources and adapt to changing analytics needs.

Cloud-based data analytics[12] presents a solution by offering elastic scalability, enabling organizations to scale resources in alignment with data growth. This addresses performance concerns and empowers organizations to process and analyze data efficiently.

However, the transition to the cloud is not without challenges. Data integration complexities, security concerns, cost management, and vendor lock-in pose significant obstacles that need to be effectively addressed to ensure a successful adoption of cloud-based BI[4].

While cloud-based BI platforms offer centralized data governance capabilities, organizations must still establish robust data management practices to ensure data quality, consistency, and compliance with regulations. This requires clear data ownership, defined data lineage, and effective data monitoring.

II. LITERATURE REVIEW

Table I. Reference To Papers

Article/Author	Year and Citation	Tools/ Software	Technique	Evaluation Parameter
Wei Han, Jie Huang, Xiaodong Zhang[3]	2010	MapReduce programmi ng model, Cloud Services	Cloud Computing, Elastic Cloud analytics	Data Security, Latency
Baek-Young Choi, Amit Sheth, Karthik Gomadam[10]	2011	Business Intelligence Tools, Cloud Services	Cloud Computing	Scalability, Collaboration
Stefan C. Müller, Christina Keller[13]	2012	Business Intelligence Tools	Cloud Computing	Accessibility, Cost Savings
Ajith Abraham, Suraj P. John, Johnson P. Thomas, Francis C. M. Lau[1]	2012	Energy Efficiency, Cloud Services	Cloud Computing	Scalability, Cost Savings, Energy Consumption, Environmental Impact
P. Velinov, S. Stojanovic[9]	2014	Business Intelligence Tools, Cloud Services	Cloud Computing	Cost- Efficiency, Accessibility
Fei Chen, Shiyong Zhang, Ke Zhang, Raheel Gillani[8]	2014	Hadoop, MapReduce programmi ng model, Cloud Services	Cloud Computing	Scalability, Cost-Effectiveness
Xiangrui Meng, Joseph Bradley, Shivaram Venkataraman, Manish Amde, Sean Owen, Doris Xin[12]	2016	Apache Spark, Hadoop	Cluster Computing	In-Memory Processing, Speed
Dimitrios Koutsomitropo ulos, Christos Doulkeridis, Kjetil Nørvåg[14]	2017	Big Data Platforms, Cloud Services	Cloud Computing	Scalability, Flexibility
Shuai Zhang, Yong Zhong, Han Su, Qiao Liu, Jianga Shang[11]	2018	Hadoop, Spark, Cloud Services	Cloud Computing	Scalability, Real-time Processing
Sami Yangui, Reda Alhajj, Rami Yassine[15]	2018	Apache Spark, AWS	Machine Learning	Real-time Analytics, Predictive Analytics

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

III. EXISTING SYSTEM

Existing systems that relate to the concerned project of currently concerned with handling, management, storage and analysis of structured data only that is chiefly based on structured databases and files. These are the CSV (Comma- Separated Values) files that are easy to operate upon and thereby the results based on the unstructured text documents, video files, audio files, mobile activity, social media posts, satellite imagery, surveillance imagery data that is mainlydifficult to analyse[2].

Previous applications deal with usage of cost-effective cloud solutions to provide an easy solution using the predefined and inbuilt algorithms that are more time consuming. This increases the downtime of providing the insights into the organization data and hinders the speed of operations required to take action on the results generated.

The more complex the system, the more difficult it is for the personnel of the business corporation to provide the favourable outcomes. The application thus, working solely on the cloud, has to be locally downloaded on the system and thereby increasing the network traffic to process the large chunk of data. APIs (Application Programming Interface) have to be extensively used to relay the information seamlessly across various tools and a separate group dealing with its managementadds to requirements.

The solutions proposed for business needs are verified through expensive service charges of the Cloud Service Providers (CSP) putting pressure on the capital invested to generate insights on the collected information. Flexibility andscalability are compromised in easy solutions that leads to needof such costly solutions that could handle the changing needs.

Therefore, a demand for one cost effective solution to such problem of business intelligence carved out, that could rely on verified analysis plus merging the concepts of encapsulation and abstraction to achieve the necessary goals. Checking thesefactors of storage, cost, scalability, downtime, flexibility, security and latency an all-round solution would be successful.

IV. PROPOSED SYSTEM

The system we propose is to provide data operations of handling, retrieval, management and analysis for both structured and unstructured data. The data may always vary inaccordance with need to provide it in tabular, xml, CSV, text based, monitoring data, metadata, profile data, and other file formats and types. Analysis on such differentiating forms would give hidden insights that can help businesses to modifyand evolve as per the shift in trends[11].

A web application that can be used on the go through the internet would save the organization requirements for CPU andRAM that would save on expenses. The user would not be required to locally download the system on his/her own system, Cloud based Data Analytics for Business Intelligence are instead the process is much more simplified where the operations would be performed automatically. Data needs to be uploaded directly into the cloud hosted application. The working is overall simplified enabling user-friendly atmosphere.

Python merged with the hosting on a cloud platform gives a scalable and flexible application. The key is to handle the operations smoothly through the efficient downtime and solutions provided by the cloud to handle dynamic demands. It is an overall cost-effective solution which gains confidence byalso providing sustainable outcomes. The business intelligencelies in the code of the application filtering out the disadvantageous information and analysing on the related data.It provides real-time data analytics saving on the time required.Firms could generate readily available solutions from the data insights. Security of personal information is the end user chiefconcern which the corporations need to address and also adhereto the rules and regulations. The application itself does not holdany data and nor does the cloud on which it is hosted. The storage is ultimately handled by the company itself that can securely store information in its own personal cloud or server systems.

Hence, our solution in this project “Cloud based Data Analyticsfor Business Intelligence” is to build on the needs of thebusinesses and provide the supply of insights by analysing anyform of data provided. We are determined to build simple, costeffective, real-time, secure, scalable and flexible solutions.

V. METHODOLOGY

This section outlines the methodology for creating aninteractive data dashboard leveraging Streamlit[5] and Plotly[6]. The objective of this research is to facilitate the effectiveexploration and presentation of data for enhanced data analytics and decision-making in diverse domains, such as business intelligence, data science, and research.

A. Installation of Streamlit and Plotly

The initial step in the development of the interactive dashboardis the installation of essential software components. The following commands are executed to acquire the necessary libraries:

pip install streamlitpip install plotly

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

B. Development of the Dashboard Application

The interactive dashboard is implemented as a Python script. A new Python file is created to serve as the foundation for the application. The requisite libraries, Streamlit and Plotly, are imported as follows:

```
import streamlit as st
import plotly.graph_objects as go
```

C. Data Loading

In the research context, the data used for the interactive dashboard is pivotal. Typically, data is loaded from external sources. For instance, a CSV file may be used, and Pandas, a data manipulation library, is employed to import the data. The following code demonstrates data loading:

```
import pandas as pd
# Load the CSV file into a Pandas DataFramedf = pd.read_csv('data.csv')
```

D. Designing the Dashboard Layout

Streamlit, a user-friendly framework, facilitates the design and layout of the interactive dashboard. This section aims to create an intuitive user interface and define the structure of the dashboard. Various Streamlit components are employed to add titles, subtitles, and other informational elements. For instance:

```
# Set the title of the dashboardst.title('My Dashboard')
# Add a subtitle
st.subheader('This is a simple dashboard.')
```

E. Integration of Plotly Visualizations

Plotly is a powerful library for creating interactive visualizations. In this step, Plotly charts are embedded within the Streamlit application to visualize data effectively. Thest.plotly_chart() function is employed to achieve this integration:

```
# Create a Plotly bar chart
fig = go.Figure(data=[go.Bar(x=df['x'], y=df['y'])])# Display the bar chart in the dashboard st.plotly_chart(fig)
```

F. Running the Streamlit Application

The final step is executing the Streamlit application, which launches the interactive dashboard in a web browser. The following command is used to run the application:

```
python dashboard.py
```

This command initiates the Streamlit server, and the dashboard is accessible at http://localhost:8501. It allows users to interact with the data and visualizations presented within the dashboard.

G. Deployment Considerations

While not an explicit part of the methodology, it is important to acknowledge that the interactive dashboard can also be deployed to a production environment to make it accessible to a broader audience. Various deployment options, such as Heroku, AWS Elastic Beanstalk, and others, can be explored to host the dashboard.

In summary, the methodology presented herein provides a systematic approach for creating an interactive data dashboard using Streamlit and Plotly. It encompasses installation, data loading, design, integration of visualizations, and the practical steps for running and deploying the dashboard, ultimately enhancing data exploration and analysis capabilities.

VI. COMPONENT EXPLANATION AND IMPLEMENTATION PLAN

A. Streamlit

- st.title(): Sets the title of the dashboard
- st.subheader(): Sets the subtitle of the dashboard
- st.sidebar(): Creates a sidebar
- st.write(): Displays text, images, and other components in the dashboard
- st.plotly_chart(): Displays a Plotly visualization in the dashboard

B. Plotly

- go.Figure(): Creates a Plotly figure
- go.Bar(): Creates a bar chart
- go.Line(): Creates a line chart
- go.Scatter(): Creates a scatter plot

C. Additional Components

- st.selectbox(): Creates a select box that allows users to select a value from a list
- st.button(): Creates a button that triggers an action
- st.table(): Displays a table in the dashboard.

D. Environment Setup

- Choose a cloud platform (e.g., AWS, Azure, Google Cloud) based on your organization's preferences and requirements. We have selected AWS cloud platform.
- Select the appropriate cloud service tiers, instancetypes, and configurations for data storage, processing,and analytics.

E. Data Generation and Selection

- Generate or gather representative datasets that mirror the data complexity and diversity of the organization's actual data sources.
- Ensure data includes structured and unstructured information to simulate real-world scenarios.

F. Cloud Services Configuration

- Set up cloud-based data storage solutions such as data warehouses, databases, and object storage.
- Configure networking settings, security groups, access controls, and encryption mechanisms for data protection.

G. Analytics Tools Selection

- Choose analytics and visualization tools (e.g., Tableau, Power BI) compatible with the selected cloud platform.
- Configure the tools for integration with the cloud datasources and APIs.

H. Experiment Design

- Define clear objectives for the experiments, such as evaluating query performance, scalability, or advanced analytics capabilities.
- Design specific scenarios that represent real-world use cases, including querying large datasets, generating reports, or running predictive models.

I. Performance Metrics

- Identify performance metrics for evaluation, such as query execution time, resource utilization, data processing speed, and visualization rendering time.

J. Data Preprocessing

- Preprocess the raw data to ensure data quality, consistency, and relevance for the experiments.
- Perform data transformations, cleaning, andaggregation as necessary.

K. Execution and Data Collection

- Execute experiments according to predefined scenarios, using realistic user queries and workloads.
- Collect relevant performance metrics and results,recording execution times and resource consumption.

L. Data Analysis

- Analyse the collected data to assess the cloud platform's performance under different scenarios.
- Compare results with baseline performance (if available) to measure improvements achieved by the cloud-based solution.

M. Scalability Testing

- Conduct scalability tests by gradually increasing the data volume and query complexity to assess the platform's ability to handle growth.

N. Advanced Analytics Evaluation

- If exploring advanced analytics (e.g., machine learning), train models on sample data and test their accuracy and predictive capabilities.

O. Visualization and Reporting

- Use analytics tools to create visualizations and reports based on the experimental results.
- Present findings in a clear and concise manner,highlighting performance gains and insights.

P. Iterative Improvement

- Use experiment results to identify areas for improvement and optimization in cloud infrastructure, data processing, and analytics workflows.

Q. Documentation

- Document the experimental setup, procedures, datasets used, performance metrics, results, and observations.
- Provide detailed explanations to ensure reproducibility and transparency.

By adhering to the given components and software tools that have been listed the application is successfully run on python and cloud environment. The given steps entail the necessary details citing how and in what manner the application can be successfully operated. Businesses when using this application can produce substantial results that could promote their growth. Based on the ease of use, numerous filter options and segmentation of outcomes according to need, it proves to be a balanced software application.

The results of this application to provide analytical information for businesses are provided via the visualizations generated. Entering a raw dataset and selecting the appropriate filters andthe fields to compare would thereby produce the expected outcomes. Some examples of the outputs produced on a given data are given as follows:

In Fig. 1 we have shown the bar graph and pie chart representation of sales over two fields in category wise sales and region wise sales depicting various percentage on one fieldof the data.

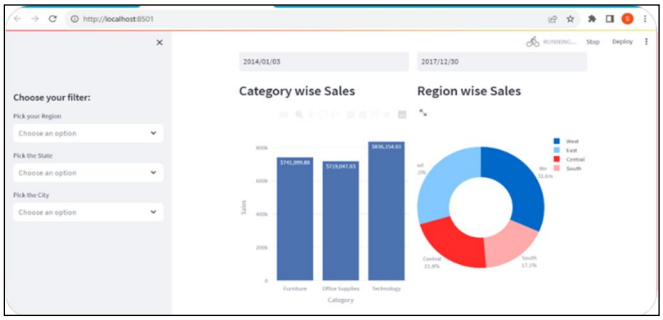


Fig 1. Category and Region Wise Bar Graph and Pie Chart

In Fig. 2 we have shown the Time Series analysis visualizationin the form of a line graph where the y-axis amount generated is matched over the x-axis month label.

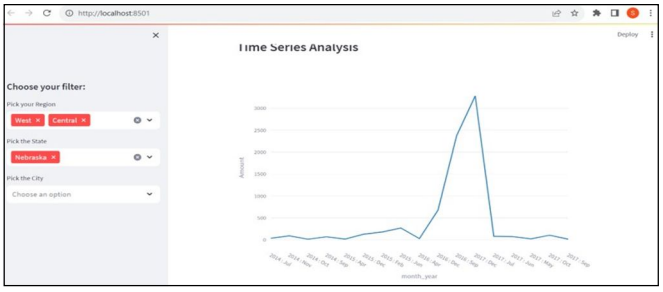


Fig. 2. Time Series Analysis

In Fig. 3 we have shown another filtered output having the same filter inclusions above but this time it depicts 2 pie chart analysis diagrams.

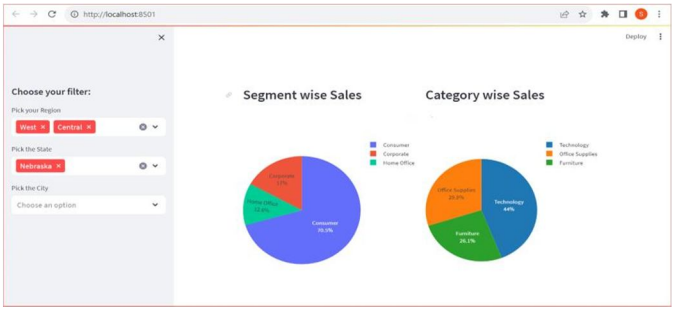


Fig. 3. Chart Analysis

The journey from traditional on-premises business intelligence to cloud-based data analytics represents a transformative shift that promises organizations unprecedented scalability, agility, and insights. This transition addresses the limitations of conventional systems and unlocks a new realm of possibilities for data-driven decision-making.

In conclusion, the adoption of cloud-based data analytics for business intelligence presents a paradigm shift with profound implications:

1) *Agility and Scalability:* Cloud platforms offer the flexibility to scale resources on-demand, accommodating growing data volumes and varying workloads. This scalability ensures optimal performance without the constraints of physical hardware.
2) *Accessibility and Collaboration:* Cloud-based solutions enable remote access, fostering collaboration among geographically dispersed teams. Real-time data sharing and interactive dashboards empower stakeholders to make informed decisions collaboratively.
3) *Advanced Insights:* Integration of machine learning and AI enables organizations to predict trends, identify anomalies, and extract deeper insights from data. This fuels innovation and proactive decision-making.
4) *Cost Efficiency:* Cloud services operate on a pay-as-you-go model, eliminating the need for upfront investments in hardware. Organizations can optimize costs by provisioning resources based on actual usage.
5) *Future Readiness:* Cloud-based analytics platforms are well-equipped to handle future trends, such as big data, IoT, and evolving analytics techniques, positioning organizations to remain competitive.
6) *Data Security and Compliance:* Implementing robust security measures on the cloud safeguards sensitive data, and compliance frameworks ensure adherence to regulations.
7) *Enhanced User Experience:* Cloud-based BI platforms offer intuitive user interfaces and self-service capabilities, empowering users to analyze data independently without relying on IT expertise.
8) *Accelerated Insights to Action:* Cloud solutions enable organizations to quickly transform raw data into actionable insights, driving faster and more informed decision-making.
9) *Improved Data Governance:* Cloud-based BI platforms facilitate centralized data governance, ensuring data quality, consistency, and accessibility across the organization.
10) *Reduced IT Burden:* Cloud providers manage the underlying infrastructure and maintenance, freeing up IT teams to focus on strategic initiatives.
11) *Enabling Data Democratization:* Cloud-based BI platforms promote data democratization by providing access to insights for a wider range of users, fostering a data-driven culture.

VII. CONCLUSION AND FUTUREWORKS

However, this transition is not without challenges. Data integration complexities, security concerns, cost management, and potential vendor lock-in require careful consideration and strategic planning. A successful implementation demands collaboration across IT, analytics, and business units, along with a continuous focus on optimizing performance and aligning analytics with business goals.

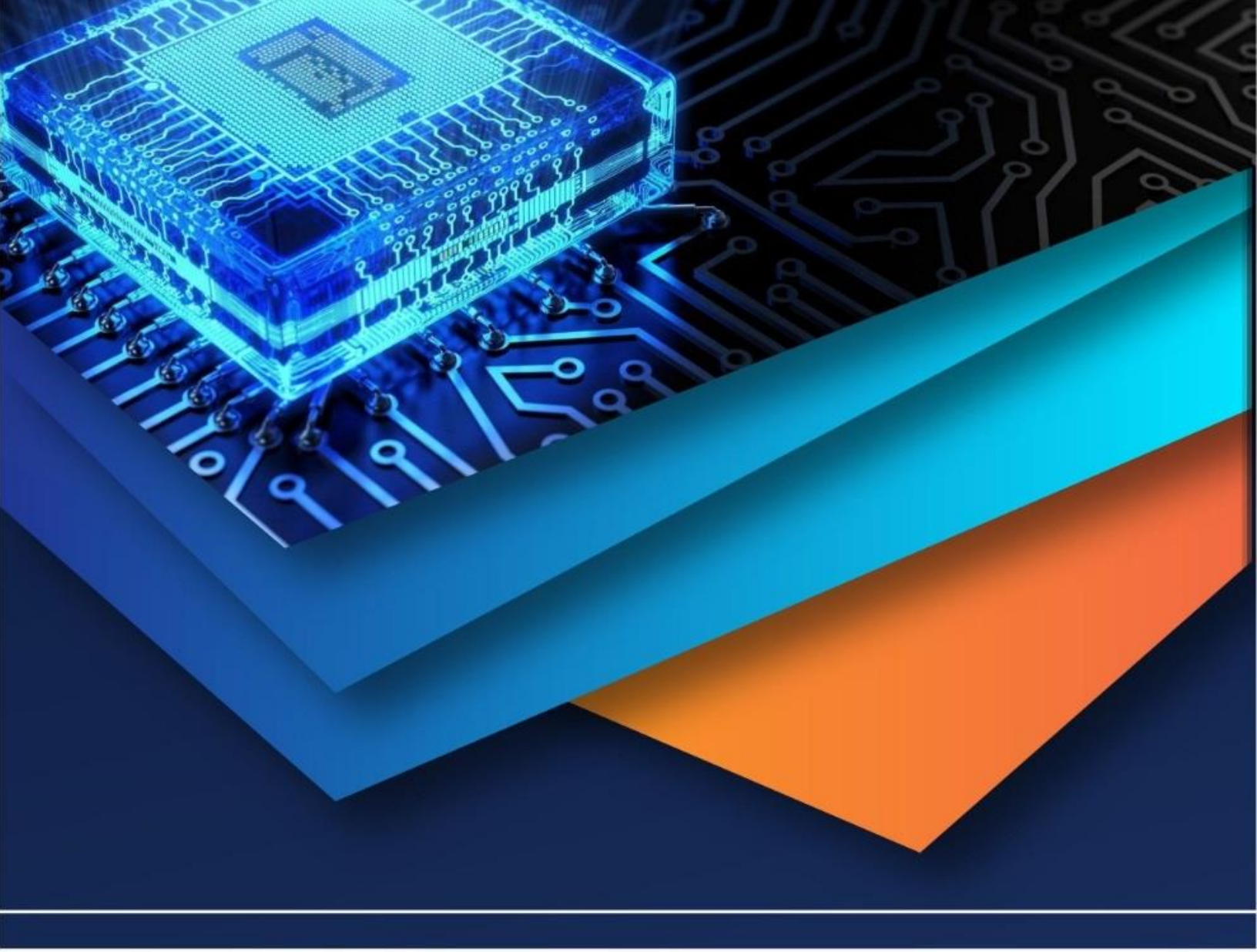
As technology continues to evolve, the cloud-based data analytics landscape will evolve too. Staying abreast of emerging trends and innovations will be crucial for organizations to remain at the forefront of data-driven strategies. The transition to cloud-based data analytics for business intelligence is not merely a technology shift; it's a transformative journey that empowers organizations to harness data's full potential, enabling them to navigate the complexities of the modern business landscape with confidence.

VIII. ACKNOWLEDGMENT

We feel ourselves remarkably privileged to have earned the final milestone during the course of our project because the success and ultimate result of our project required a great deal of support and aid from many individuals. We would like to thank them for their aid and direction, without which we wouldnot have been able to accomplish as much as we have. The authors are truly grateful to Ms. Sukhmeet Kaur, teacher in charge, Apex Institute of Technology, Chandigarh University, Punjab, India, for her full support extended to carry out this research.

REFERENCES

[1] Chen, H., Chiang, R. H. L., & Storey, V. C., "Business intelligence andanalytics: From big data to big impact", MIS Quarterly, pp. 1165-1188,2012
[2] Davenport, T. H., Harris, J. G., & Shapiro, J., "Competing on analytics:
The new science of winning", Harvard Business Press, 2010
[3] Eckerson, W., "Performance Dashboards: Measuring, Monitoring, andManaging Your Business", Wiley, 2009
[4] Elbashir, M. Z., Collier, P. A., & Sutton, S. G., "The role of organizational absorptive capacity in strategic use of business intelligence to support integrated management control systems", TheAccounting Review, 86(1), 155-184, 2011
[5] Kimball, R., & Ross, M., "The Data Warehouse Toolkit: The DefinitiveGuide to Dimensional Modelling", Wiley, 2013
[6] Marz, N., & Warren, J., "Big data: Principles and best practices ofscalable real-time data systems", Manning Publications, 2015
[7] Mittal, M. L., & Mittal, A., "An overview of cloud computing and itsimpact on business intelligence in healthcare", Procedia Computer Science, 31, 208-213, 2014
[8] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V., "Criticalanalysis of big data challenges and analytical methods", Journal of Business Research, 70, 263-286, 2017
[9] Xiangrui Meng, Joseph Bradley, Shivaram Venkataraman, Manish Amde, Sean Owen, Doris Xin, "MLlib: Machine Learning in Apache Spark", Journal of Machine Learning Research, pp 1-7, 2018
[10] Zhu, K., & Xu, S. X., "The complementarity of information technology infrastructure and e-commerce capability: A resource- based assessment of their business value", Journal of ManagementInformation Systems, 29(1), pp. 167-200, 2012



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY