

# Exploratory Data Analysis in Cloud Computing

**Authers' name: Shreya Ramachandra**

## **I. ABSTRACT**

By utilizing the distributed computing capabilities of the cloud data processing engine, data analysis services in the cloud enable users to study and explore large-scale in-cloud datasets fast and interactively with a straightforward graphical user interface. Users may build a data flow graph in a GUI and examine the outputs that are shown using visualization tools.

## **II. INTRODUCTION**

Exploratory data analysis (EDA) is a crucial step in the data science process that allows for the discovery of patterns, outliers, and other insights in a dataset. With the increasing popularity and accessibility of cloud computing, EDA can now be performed on large datasets without the need for costly and time-consuming data storage and processing. This paper aims to explore the benefits and limitations of performing EDA in cloud computing environments, as well as to provide an overview of popular tools and techniques for EDA in the cloud.

## **III. METHODS**

This paper collects the information through a literature review of academic and industry articles, as well as through the examination of popular cloud-based EDA tools and techniques. The literature review focused on the benefits and limitations of performing EDA in cloud computing environments, as well as on the most used cloud-based EDA tools and techniques.

## **IV. RESULTS**

Performing EDA in cloud computing environments offers several benefits, including the ability to work

with large datasets without the need for costly and time-consuming data storage and processing, the ability to collaborate with other data scientists in real-time, and the ability to easily scale up or down as needed. Additionally, cloud-based EDA tools and techniques have become increasingly advanced, allowing for the performance of complex analysis and visualization.

However, there are also limitations to performing EDA in cloud computing environments, such as data security and privacy concerns, as well as the potential for vendor lock-in. Furthermore, cloud-based EDA tools may lack some of the advanced features and functionalities of traditional on-premises tools.

Popular cloud-based EDA tools include Amazon's QuickSight and Google's Data Studio, while popular cloud-based EDA techniques include data visualization, data sampling, and data cleaning.

## **V. CONCLUSION**

Performing EDA in cloud computing environments offers several benefits, including the ability to work with large datasets without the need for costly and time-consuming data storage and processing, the ability to collaborate with other data scientists in real-time, and the ability to easily scale up or down as needed. However, there are also limitations to performing EDA in cloud computing environments, such as data security and privacy concerns, as well as the potential for vendor lock-in.

Furthermore, cloud-based EDA tools may lack some of the advanced features and functionalities of traditional on-premises tools. Overall, while cloud-based EDA has many advantages, it is important to weigh the benefits against the limitations and select the right tools and techniques for the specific needs of the analysis.

## REFERENCES

- [1] “1.1.1. What is EDA?,” *Nist.gov*. [Online]. Available: <https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm>. [Accessed: 04-Jan-2023].
- [2] A. Venkatachar, “EDA in the cloud,” *From Silicon To Software*, 26-Aug-2021. [Online]. Available: <https://blogs.synopsys.com/from-silicon-to-software/2021/08/26/cloud-based-eda-tools/>. [Accessed: 13-Jan-2023].
- [3] D. Mathapati, “Top 5 AWS data analytics services to master in 2022,” *CloudThat Resources*, 30-Aug-2022. [Online]. Available: <https://www.cloudthat.com/resources/blog/top-5-aws-data-analytics-services-to-master-in-2022>. [Accessed: 24-Jan-2023].