Hand Gesture Recognition Using Machine Learning with Convolutional Neural Network (CNN)

Prathamesh Kunjeer[1], Madhura Khedkar[2], Mansi Chavan[3], Niyukti Wani[4], Prof W. S. Rahane[5]
Department of Information Technology, NBN Sinhgad School of Engineering, Ambegaon BK., Pune[1,2,3,4,5]

Abstract: *In this era of technology, the most valued asset can be 'Data'. With the increasing number of data, the value of it keeps increasing. We have started to store and manipulate data to achieve some particular goals or business requirements but with the increasing number of data, storing it has become a complex and tedious task. With the use of some advanced technologies like Hadoop, we simplified the data storing process but due to rapid development and excessive use of AI and ML tons of data is collected. For this, the paper provides us with an effective solution to store data over the cloud with numerous benefits over traditional data storage methods by developing a data lake using AWS. Furthermore, the functionalities of Data Lake include managing and storing sorted as well as unsorted data, gathering various analytics from the data lake as per business requirements.*

Keywords: Data Storage Techniques, Data Lake, AWS Data Storage, AWS Solutions, Amazon S3, etc.

I. INTRODUCTION

To introduce the concept of Data-lake [1] let us understand a following example. Imagine a data lake as a huge virtual cloud storage where 'n' number of users can upload / download data and later use it for processing to extract useful results or conclusions. Talking more about the traditional methods for storing data i.e., data warehouses they work fine until the data is predefined and acute. However, with the modernization in technology and excessive use of Artificial Intelligence and Machine Learning, the number of raw and unfiltered data is increasing [2]. To cope up with this situation data lakes prove to be an easy and efficient means of data storage. Major issues such as frequent data loss, low data quality, high data storage cost is addressed in data lakes. While the data keeps generating, organizations need a way to use their data part from just storing into an effective tool for enhancing their businesses. Data in S3 data lake has a data longevity rate of 99.9999999 %, which puts it ahead of other competitors [1] This paper helps us to answer questions like is storing data a costly and complex thing? How can one establish a secure data lake environment ready for business use? How can one integrate existing data warehouse with upcoming data lake technologies?

II. LITERATURE REVIEW

The authors [1] give us a descriptive idea about what is a data lake, its major characteristics, with some information of existing data storage systems like Hadoop, AWS, and Azure etc. Hadoop is one of the traditional data storing methods uses Map Reduce for data analysis. On the other hand, AWS provides exciting features like enhanced data privacy, smooth data integration and user friendliness. The authors [2] compares the functionalities of data lake and data warehouses. We get a brief idea about the existing data storing methods with all the positives and negatives.

Our major focus will be to study and import a large amount of data that can come in real time, scale data of any size while saving time in defining data structures, schema and transformations, effective and affordable approach to store, manage and analyze data to improve performance of the applications and meet all possible requirements of the user. When the amount of data generated was low, traditionally data was stored on physical

drives and magnetic disks. Users interacted with the data storage mechanics physically when they require any data in real time. Users need to predefine some data structures for storing the data, think about efficient memory management techniques, efficient algorithms, and process the data before entering the storage [3].

Later querying the data is a slow process with increasing number of data the time required for fetching the data from the storage increases. Some of the major drawbacks of traditional system are:

- Physical devices are required for storage.
- User must run maintenance tools manually.
- High initial investment and effort.
- More prone to cyber threats like virus attacks

III. PROPOSED SYSTEM
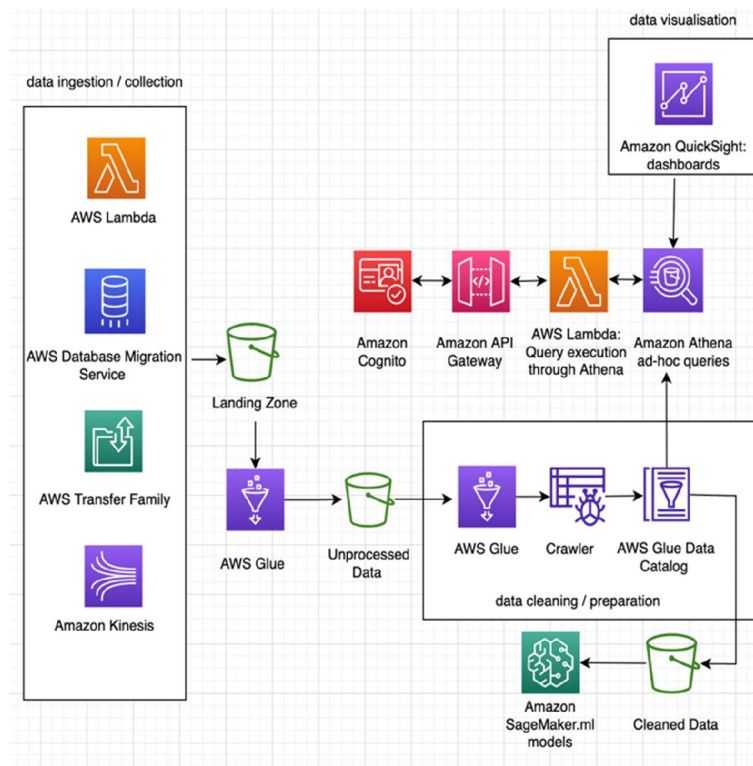
Figure 1: Proposed system block diagram

Data warehouse v/s Data lakes

Data warehouses are traditional data storage methods whereas data lakes are a modern updated version of existing data warehouses. Major benefits of data-lake over data-warehouses are:

- Democratize data
- Improve data quality
- Higher Scalability and Versatility
- Schema Flexibility
- Advanced Analytics

Data storage in native format

Below table depicts the major differences between data lake and data warehouses [1].

Sr. No	Dimensions	Data Warehouses	Data Lakes
1]	Data can be stored in		
	• structured format	✓	✗
	• unstructured format	✗	✓
	• raw format	✗	✓
	• processed format	✓	✗
2]	Schema		
	• on-write	✓	✗
	• on-read	✗	✓
3]	Scalability		
	• Volume	Large	Extremely large
	• Cost	Moderate	Low
4]	Architectural Design		
	• Hierarchical	✓	✗
	• Flat	✗	✓
5]	Design Complexity		
	• Joins	✓	✗
	• Processing	✗	✓
6]	Efficiency		
	• CPU/IO	User efficient	Moderately user efficient

IV. METHDOLOGY

Developing a data lake by selecting an appropriate application for building a data lake, we need to design the data lake, which will follow the given steps:

1. Mapping incoming data and user requirements
2. Create IAM users – IAM user is an entity you can create in AWS to represent the person or application interacting with the Data Lake. Create IAM users as per needs in a particular location. Grant appropriate permissions by attaching existing / creating security policies.
3. Create IAM roles – IAM role is an entity, which has the permissions to make service requests to Data Lake.
4. Create S3 bucket for storing data.
5. Configure the data lake – Configuring the data lake includes creating and managing administrators, creating a database, registering your S3 bucket as your data lake's storage.
6. Configure and Run Crawler – Every data in the lake should be cataloged. The catalog is automated using crawlers in AWS Glue. Create an appropriate crawler as per business needs grant permissions accordingly.
7. Granting user permissions to the catalog.
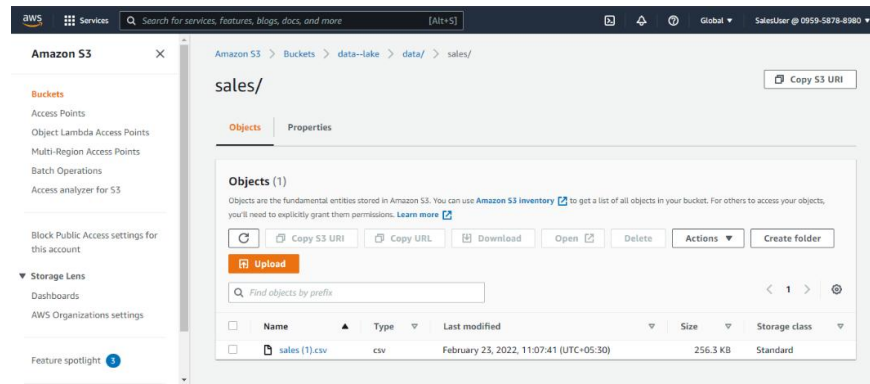8. Query data from the data lake as per business needs.

V. RESULT AND DISCUSSION

Figure 2: Sales User View of Data Lake

The sales user uploads all the product specifications and details on the data lake for the customers to go through.
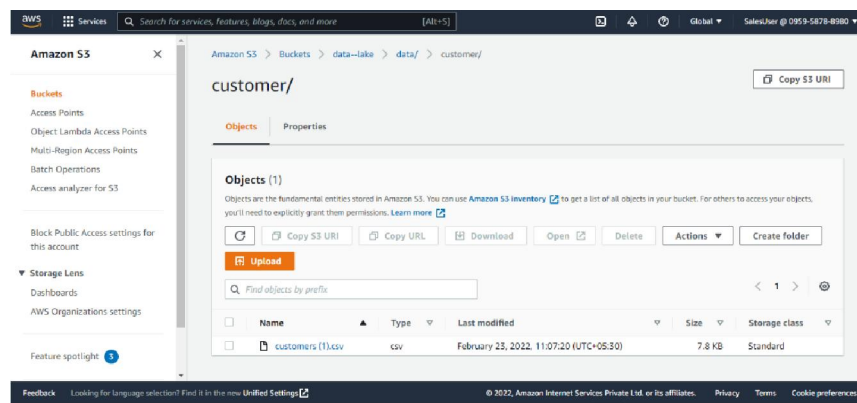

Figure 3: Customer User View of Data Lake

The customers go through various products with their specifications and decides whether to purchase or not.
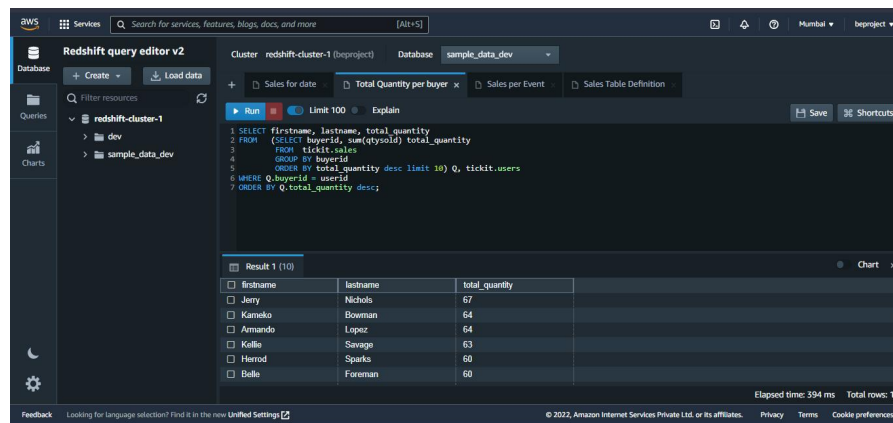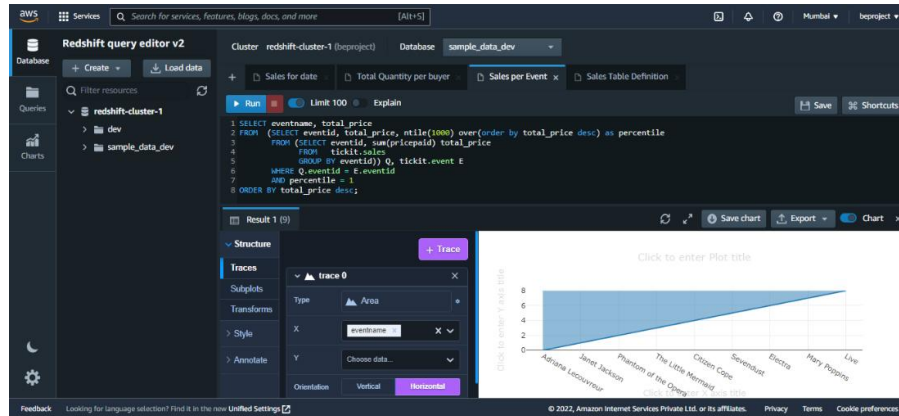

Figure 4: Querying Data from the Data Lake

The data lake admin queries the required data which is requested by any particular user as per his business needs.

Figure 5: Area Graph for Sales per Event

Later the data lake administrator can draw various analytical conclusions based on the fetched data from the lake.

VI. CONCLUSION / FUTURE WORK

With increasing amount of data from numerous sources, developing a Data Lake is a need, to store large amount of data efficiently that comes in real life. Security and user accessibility plays a major role nowadays. AWS offers excellent and efficient end-to-end framework with solutions like security, managing, monitoring data at low cost effectively. Apart from being a widely use technology, AWS also has various integrity features to upscale existing data warehouses into data lakes at low cost without any major complexities

AWS provides the broadest selection of analytics services that fit all your data analytics needs and enables organizations of all sizes and industries to reinvent their business with data. From data movement, data storage, data lakes, big data analytics, log analytics, streaming analytics, and machine learning (ML) to anything in between, AWS offers purpose-built services that provide the best price-performance, scalability, and lowest cost and mostly used in store large amount of data is structured or unstructured.

REFERENCES

[1] Aakash Aundhkar, Shweta Guja, A review on Enterprise Data Lake Solutions, Journal of Science and Technology, Volume 06, Issue :01| August 2021.

[2] Tanmay Sanjay Hukkeri, Vanshika Kanoria, Jyoti Shetty, A study of Enterprise Data Lake Solutions, International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 05| May 2020.

[3] Amra Munshi, Yasser Abdel-Rady I Mohamed, Data Lake Lambda Architecture for Smart grids big data analytics, IEEE Issue: 23 July.

[4] Bozena M-M, Marek S, Dariusz M. Soft and decarative fishing of information in Big Data Lake. IEEE Transactions on Fuzzy Systems, 2018, 1(99):1-6.

[5] A. Cravero, O. Saldana, R. Espinosa, and C. Antileo, "Big data architecture for water resources management: A systematic mapping study," IEEE Lat. Am. Trans., vol. 16, no. 3, pp. 902-- 908, 2018.

[6] Sophia Boing Righetto, Eduardo Luiz Martins, Andre Luiz Pereria, Data Lake Architecture for Distribution System Operator, 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) | 978-1-7281-8897-3/21/$31.00©2021IEEE|DOI: 10.1109/ISGT49243.2021.9372181

[7] ByungRai Cha, Jong won Kim, Design and Implementation of connected data lake for a reliable data transmission.

[8] Tanmay Sanjay Hukkeri, Vanshika Kanoria, Jyoti Shetty, A study of Enterprise Data Lake Solutions, International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 05|May 2020

[9] Yi-Hua Chen, Hsin-Hsin Chen, and Po-Chun Huang, Enhancing the Data Privacy for Public Data Lakes, Proceedings of IEEE International Conference on Applied System Innovation 2018

[10] J. Sawadogo, Pegdwende and Darmont, "On data lake architectures and metadata management," J. Intell. Inf. Syst. Springer, pp. 1--24, 2020.

[11] Mukund Rajeshwar, Rajesh Bharati, "Function as a Service in Cloud Computing: A survey", International Journal of Future Generation Communication and Networking Vol. 13, No. 3, (2020), pp. 3291–3297.