

Cloud based Business Intelligence and Sales Forecasting using AWS

Ankit Kumar Nath, Department of Computer Science, GIET University,
ankitkumarnath121@gmail.com

Omshree Patra, Department of Computer Science, GIET University,
omshreepatra@zohomail.in

Abstract: The use of cloud-based solutions for Business Intelligence (BI) integrated with sales forecasting models oriented towards the use of Amazon Web Services (AWS) is discussed. Various services offered by AWS like Amazon S3, AWS Glue, Amazon SageMaker, Amazon Redshift and Amazon QuickSight are discussed where scalable solutions can be availed for processing, analysing and visualising business data. The service usage by organizations is discussed for building accurate forecasting models and creating actionable items related to the business decisions. The salient features of the AWS infrastructure, internal working practices and overall cost-effective measures are discussed in detail showing the potential of the cloud in BI and sales forecasting.

Keywords - Cloud Computing, Business Intelligence, Sales Forecasting, AWS, S3, Glue, SageMaker, Redshift, QuickSight, Cost-Effectiveness

1. Introduction

Business Intelligence (BI) refers to the technologies, tools, and processes which assist organizations to collect, analyze, and visualize business data in order to make better decisions. BI was previously located within the company and presented all the issues brought by an intricate and expensive infrastructure. With the emergence of cloud computing, BI has been transformed with a new ability, affordable price, and flexibility. AWS is a major cloud service provider and provides an entire package of integrated tools to support tasks of BI and data analytics. In this paper, the attention is paid to integrating the selected AWS solutions (Amazon S3, AWS Glue, SageMaker, Redshift, and QuickSight) that can create a cloud BI application to assist in sales forecasting. This application will be helpful to business managers to process, analyze, and visualize data and guarantee efficient and effective operation costs and scalability.

The research paper is devoted to the design and description of the cloud-based Business Intelligence and sales forecasting system, which is built on the services of AWS. It analyses the entire pipeline, including data administration in Amazon S3, information transformation with AWS Glue, prediction through Amazon SageMaker, information warehousing with Redshift and visualization of the data with QuickSight. The main focus of the present study is to develop a cloud-based Business Intelligence architecture using AWS services such as S3, Glue, SageMaker, Redshift, and QuickSight. Understanding its internal working of major AWS services involved in data storage, processing, ML model training, and visualization. Analysis of forecasting accuracy using machine learning models built in Amazon SageMaker. In addition to development of BI architecture, we evaluate its cost-effectiveness, optimization and data-driven decision-making is also investigated for marketing, inventory, and sales planning.

2. Literature Review

Smith and Johnson (2022) examined the role of cloud computing in transforming Business Intelligence by enhancing scalability, flexibility and cost-efficiency in the data-driven organizations [1]. They studied the shortcomings of conventional on-premise BI systems and

pointed to how cloud computing and, specifically, the AWS, make ease data storage, processing, and visualization. They highlighted that Amazon S3, Redshift and QuickSight services decrease the infrastructure overhead and enable advanced analytics workloads. Their analysis has found out that cloud-based BI is now critical in the contemporary enterprises that desire real-time scalable insights [1]. This is also in line with the current research which focuses on AWS-enabled BI structures that are specific to sales forecasting.

Application of machine learning in cloud-based BI set-ups was reported has shown that cloud systems allow the effective training of models and their deployment without the need of expensive on-premise systems [2]. Results showed that the services such as Amazon SageMaker are very important in reducing the threshold to implementing predictive analytics with automated training deployments and distributed computing [2]. The research by Patel positively correlates with the application of SageMaker in this current study to create time-series forecasting models that can be used to forecast sales trends and patterns, customer behavioral trends, and trend of operations [2].

Another 2020 study released as EAI series also studied the construction of cloud-based demand forecasting systems, especially to small and medium-sized businesses. The authors created an example of forecasting pipeline on cloud infrastructure, whereby, retail and inventory data are processed and showed that deploying the infrastructure on clouds lowers operational costs, and forecasting reliability is also enhanced. Their results supported the notion that small organizations that have no particular data science teams find cloud tools to be the most appropriate. This study offers a viable research study to support the usage of AWS Glue and S3 in low-cost ETL and storage procedures, which further supports the cost-effectiveness aspect of the current paper. BI dashboards also reported to be used to enhance business decision-making with the help of machine-learning models [2, 3]. Their findings showed that organizations can have an advantage when the analytical models are linked with the user-friendly dashboards, which is the core idea of this research to provide sales insights to the end-users through Amazon QuickSight. Comparing of the three providers, AWS, Google Cloud, and Azure in terms of their performance, cost, and integration was reported and analysed [3]. According to their findings, AWS indicates better end-to-end integration of the storage services with ETL services, machine learning services, and BI services, which means that it is suitable to the enterprise-wide analytics workflow [3]. Although their work has made a general comparison, it identified a gap in the comprehension of particular case studies that can be used to illustrate full pipeline forecasting that has been deployed on one cloud platform. This restricts the research gap that proves to be filled by the current study by formulating and testing a complete AWS pipeline using S3, Glue, SageMaker, Redshift, and QuickSight to be able to forecast sales.

3. Results and Discussion

In order to introduce a good BI and sales forecasting system, a strong cloud architecture is needed. AWS offers a number of services, which assist every single step of data processing, analysis, and visualization.

A. Amazon S3 (Simple Storage Service)

Amazon S3 is a scalable data storage service. It helps companies to store extensive data in structured and unstructured formats like sales records, client information and product information. The racetrack on which S3 operates is a carrier loop which is highly available and cost-effective and its billing is based on a pay-for-what-you-need move (depending on the extent of data stored). Data in the system is stored in buckets, which act as logically organized containers for managing objects. Each object is identified by a unique key, enabling simple and efficient data retrieval. The system also supports multiple storage classes, such as Standard and Intelligent-Tiering, allowing optimized cost management based on access frequency and storage requirements.

B. AWS Glue

AWS Glue is a data integration service where ready and transformative data is prepared to be utilized in analytics. It automated data extraction, transformation, and loading (ETL) operations, which simplifies the movement of data in S3 to Amazon Redshift to analyze it. AWS Glue crawls the stored data to locate and index datasets in Amazon S3, creating metadata that can be easily integrated with other AWS services. It automates ETL (Extract, Transform, Load) processes using Python or Scala scripts that run on a serverless platform. Additionally, AWS Glue supports both real-time data processing and batch jobs, enabling organizations to clean, transform, and prepare data efficiently for further analysis.

C. Amazon SageMaker

Amazon SageMaker is a full-service machine learning service enabling data scientists to create, train, and integrate traits of predictive models. SageMaker may be applied to create time series models and machine learning algorithms that make future sales predictions using historical data in sales forecasting. Amazon SageMaker provides easy-to-use built-in algorithms, ready-to-run development environments, and scalable training resources to support efficient model creation. It integrates with Amazon S3 for data storage and AWS Glue for data transformation, enabling a seamless data processing pipeline. After the model is trained, it can be deployed through SageMaker endpoints to perform real-time inference on incoming data.

D. Amazon Redshift

Amazon Redshift is a full-fledged data warehouse solution that enables businesses to query large volumes of data using complicated SQL queries. It is the same platform as S3 and Glue and it can easily interact with transformed data stored on the cloud to generate efficient queries. Amazon Redshift stores data in a columnar format, which enables faster query performance compared to traditional row-based database systems. It utilizes Massively Parallel Processing (MPP) to distribute query workloads across multiple nodes, significantly improving data processing speed and efficiency. Additionally, Redshift is designed for high availability and scalability, providing features such as automatic backups and seamless integration with various business intelligence (BI) tools for advanced data analytics.

E. Amazon QuickSight

Amazon QuickSight is a BI service powered by the cloud that people use to make and share interactive dashboards, visualizations, and reports. It is directly integrated with Redshift and S3 to help the end-user view sales data and forecasts created by SageMaker. Amazon

QuickSight provides a user-friendly interface that allows users to create interactive dashboards using simple drag-and-drop features. It automatically scales according to the number of users and the volume of data, offering an elastic and efficient business intelligence solution. Additionally, QuickSight supports advanced analytical capabilities, including predictive analytics, which can be utilized for tasks such as sales forecasting and data-driven decision-making.

F. Cost-effective Solution for BI and Sales Forecasting

The cost-efficiency of the solution is also one of the main benefits of the use of AWS in BI and sales forecasting. The pay-as-you-go system allows the businesses to pay only what they consume, and thus, cloud-based BI is cheaper than the habitual on-premises systems.

Data Storage and Transformation Costs: Amazon S3 provides a low-cost and highly scalable storage solution, allowing businesses to optimize expenses by selecting appropriate storage classes based on data access patterns. AWS Glue, with its serverless architecture, eliminates the need for infrastructure provisioning and enables organizations to pay only for the ETL jobs they execute, thereby improving cost efficiency in data transformation processes.

Sales Forecasting and Machine Learning Costs: Amazon SageMaker follows a pricing model based on the resources utilized during model training and inference. Its managed service architecture allows businesses to avoid the costs associated with maintaining and managing machine learning infrastructure, thereby enabling efficient and scalable deployment of predictive models.

Data Querying and Analysis Costs: Amazon Redshift enables businesses to start with small clusters and scale them as required, providing flexibility in managing data workloads. Additionally, the option to pause clusters when they are not in use helps optimize operational costs. Amazon QuickSight follows a per-user, pay-per-session pricing model, making it a cost-effective business intelligence solution for organizations that want to scale analytics usage without significant upfront investment.

Cost Optimization Strategies: Cost efficiency can be improved by using AWS Cost Explorer to monitor and optimize the utilization of various cloud services. Organizations can also reduce expenses by using Reserved Instances for services such as Redshift and SageMaker, which provide cost savings for long-term usage. Additionally, data tiering in Amazon S3 allows businesses to store infrequently accessed data in lower-cost storage classes, helping to further optimize overall storage costs.

1. Internal Working of the Sales Forecasting Solution

The common steps in sales forecasting using the cloud will include the following:

- I. **Data Collection and Storage:** Amazon S3 stores sales data, customer interaction data, and product data. AWS Glue can automatically ingest data (CRM systems, transaction logs, etc.) in several sources.
- II. **Data Preparation and Transformation:** AWS Glue is employed to normalize and feed raw data to easily analyzed structured forms. It could include normalization, dealing with missing values and aggregation of data.

- III. **Model Training and Forecasting:** This information is then entered into Amazon SageMaker to develop and train forecasting models with such algorithms as ARIMA, XGBoost, or even deep learning. The trained model is then transferred to deliver real time predictions.
- IV. **Data Analysis and Visualization:** The predicted and processed data is updated to Amazon Redshift where it can be queried. The sales trends, the KPIs, and the forecasts are then visualized in interactive dashboards using Amazon QuickSight.
- V. **Reporting and Decision-Making:** The findings in QuickSight are communicated to business stakeholders so that they can be able to make strategic decisions on the basis of inventory, marketing, and sales.

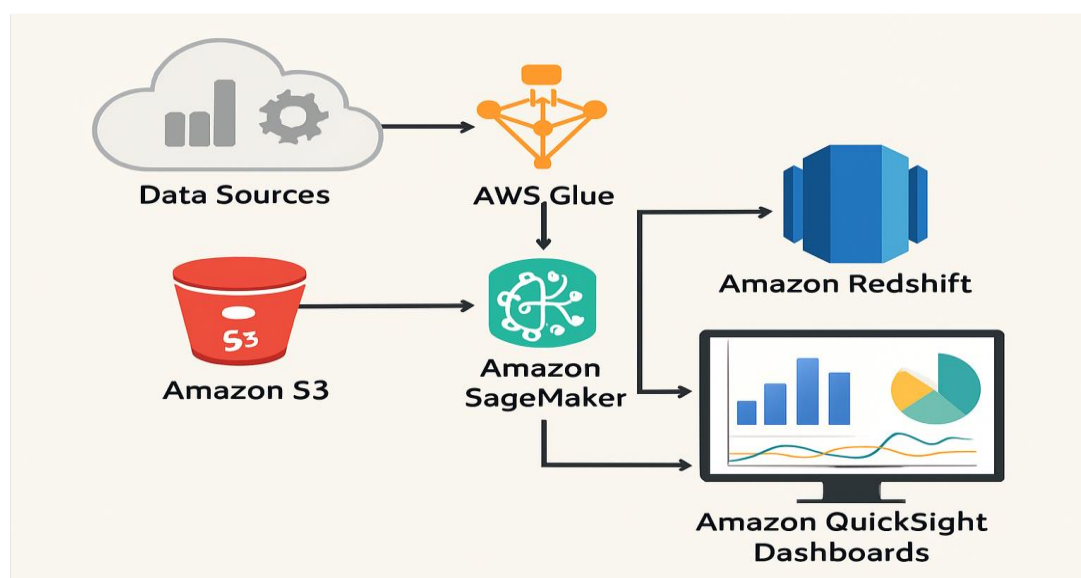


Fig. 1. Internal working of forecasting solution

4. Conclusion

In conclusion, AWS offers a powerful and affordable platform of deploying cloud Business Intelligence and sales forecast systems. With the help of such services as S3, Glue, SageMaker, Redshift, and QuickSight, companies can make correct sales predictions, perform data processing and analysis on large amounts of data, and make decisions based on data. Scalability and flexibility of AWS also mean that organizations can match with the needs of the business as they evolve besides being cost-efficient.

5. References

1. J. Smith and A. Johnson, "Cloud Computing in Business Intelligence," IEEE Transactions on Cloud Computing, vol. 7, no. 4, pp. 233-245, Dec. 2022.
2. S. K. Patel, "Data Analytics and Machine Learning in Cloud-based BI," Journal of Business Intelligence, vol. 5, no. 2, pp. 85-96, Feb. 2023.
3. AWS Whitepaper, "Amazon Web Services for Business Intelligence," Amazon Web Services, 2021. [Online]. Available: <https://aws.amazon.com/solutions/bi>.