Dynamis Hub

Abstract

The domain of this research is the analysis of two out of the three largest IoT(Internet of Things) cloud platforms that are Amazon Web Services(known as AWS) and Microsoft Azure, compare and contrast them, and finally suggest one of these based on this research. At first, we focus on their services and pricing criteria which are critical to evaluate in the process of big data as a data warehouse. Second, we evaluate their ease of use and elasticity in the services they provide. Finally, we compare their performance in services as a data warehouse according to other data warehouse providers and suggest the best option. This research will also examine the results of our process in a big data project and we will evaluate these results according to the visualization and machine learning we made in Python programming language.

Introduction

The demand for cloud computing as a data warehouse has rapidly increased in the last decade which has driven significant scaling of cloud computing platform usage by huge companies. The pricing of the biggest cloud computing companies such as Microsoft, Google, and Amazon has been changed through the years in order to be affordable for the companies. Cloud providers change the way companies save and operate with data to be easier to use and more efficient in the services that they provide at a better price. Hence, in this period of big data, these cloud services are playing an important role with strong attention to the design of the database schema according to the basic criteria of evaluation.

Amazon Web Services

Amazon Web Services was first announced in 2006, is a cloud computing platform which gives the opportunity via the services that provide to develop, operate, and manage multiple applications in databases, and in proportion to all the above the companies that use AWS decreased the money they invest for data warehousing. Furthermore, the services AWS provides are computing power, data storage for databases, and artificial intelligence which allows users to customize their services according to the necessity of every database that they use. These services with management, security, and monitoring tools supply flexibility and efficiency, allowing businesses to easily and securely benefit from their digital transformation.

Microsoft Azure

Microsoft Azure was first announced in 2008 at the Professional Developers as a project with the code name ‘Project Red Dog’ and six years later it was renamed as Azure. The first period started with limited services and without additional coding. After that Azure started as Infrastructure as a Service which was the best approach for customers to control over the cloud, they changed the operations from bottom to top from PaaS (Platform as a Service) to IaaS (Infrastructure as a Service). When the era of big data came Microsoft launched Azure Data Lake Store and Azure Data Lake Analytics to provide an end-to-end Big Data analytics platform which would later acquire the R language and the other services to supply the service providers with end-to-end connected devices such as Event Hub, IoT, SQL Database, Power Bi and others to offer a full package of analytics environment. Nowadays, the Azure platform adopts Azure Arc which allows the customers to manage their Virtual Machines (VMs). Hence, Azure is a cloud computing provider which combines many different technology services like SaaS (Software as a Service), PaaS (Platform as a Service), and IaaS (Infrastructure as a Service) and these make this service user-friendly.

Compare And Contrast AWS – Microsoft Azure

To choose one platform from these two cloud platforms AWS (Amazon Web Services) and Microsoft Azure is difficult both have many advantages and disadvantages and of course many different services. In the case of delving deeply, AWS differentiates for the wide variety of services that provides, and on the other hand, Azure gives the flexibility for the payment and the combination of the different services that can operate in the cloud. AWS is a secure and reliable platform with many operations for the developers which are appropriate for big companies. Although, AWS is not so easy to use, and it is more expensive than Azure.

Moreover, Azure is flexible and easily adaptable in pricing, Microsoft’s Platform as a Service (PaaS) is known for the accuracy and expansibility that provides to the customers. In addition to these, Microsoft provides the support for open-source code and confronts with compatibility issues that emerge by explaining how they solve them. The transfer of data is extra in the payment.

To focus more on the technical components that they have it needs to compare them in performance at scale, elasticity, ease of use, and efficiency at cost.

Performance at scale

AWS: It stands out for its exceptional performance at scale, offering extensive utilization of computational resources and high reliability.

Azure: It supplies also exceptional performance at scale with high adaptability and availability to manage big data tasks.

Elasticity

AWS: It offers a remarkable elasticity, which allows it to be easily adaptable to compute open-source applications according to the company's needs.

Azure: it also features with robust elasticity allowing for the automatic adjustment of services to meet demand.

Ease of Use:

AWS: It has remarkable documentation and clear tools but it may be complex for some users due to the complexity of the platform.

Azure: It has been effectively integrated with Microsoft products, offering easy integration for users within the Microsoft environment.

Cost Efficiency:

AWS: It is considered as cost-efficient, but users need to be careful with the operation of computational resources to avoid high costs.

Azure: It yields competitive pricing and payment flexibility which makes it financially efficient.

Summarize

To sum up, Amazon Web Services (AWS) and Microsoft Azure are made for specific occasions and needs. These platforms have some benefits and some services that every company needs to consider which one is the most suitable for their data warehouse. The parameters they need to evaluate in order to choose which platform are the following:

The wide Variety of Services:

AWS is the most suitable platform for this occasion because it has a wide variety of operations and services that other cloud computing platforms don’t have.

Economic Efficiency:

If the company needs a data warehouse which the most important role plays the economic efficiency, then Azure may be preferable due to its competitive pricing and billing flexibility.

Flexibility and Integrated Solutions:

If the company already has Microsoft products and services, then Azure provides easy compatibility and integration.

Development and Usage Environment

If the company needs a user-friendly environment and the desire for rapid development are primary concern, then Azure may be preferable because it can combine with the other Microsoft products easily.

To consider all the above, the choice depends on the context and needs of each company because every platform has advantages and disadvantages. It is recommended that a detailed assessment of requirements should be conducted before the selection of the suitable platform.

Future Concerns and Improvements

Usage simplification:

Simplification of the installation and configuration process for services.

Improvement of educational resources and documentation for easier learning and usage

Integration of Products:

Better integration of the platforms with other popular tools

Strengthening compatibility with third-party products for more integration options

Cost Management improvements

Better cost management tools to be more transparent usage of services and improvement of cost prediction and control tools.

Security and privacy enhancements

Continuously improvement of security and privacy protection.

More tools for monitoring and threat mitigation.

Advanced Technologies Support:

Expansion of technology support in case to provide more services for machine learning, blockchain, and data analytics.

These improvements will make the platforms more accessible, cost-effective, and adaptable.

Introduction to Dataset Analysis

This is the analysis of the dataset that we proceed with the data analysis steps to extract important information about the customers, sales, and territory to look at how these three components are affecting each other and how much. We used MySQL as the data warehouse, and we combined there the CSV files in case to make the database more readable and efficient the next steps we used Python as the programming language to achieve data cleaning, transformation, and exploratory data analysis, we prepare the dataset for analysis by doing all these steps. Furthermore, we used machine learning to understand the patterns and correlations between this information.

We used data visualizations in case to extract some insights, then we used statistical techniques to see if this database there were some patterns between the information that we analyzed and finally we used machine learning to prove the accuracy of my findings which we have as a goal to deeply understand the correlation between the sales and how they are affected by customers.

First of all, we took the customer distribution by geographic region and we observed that the customers are mainly in the country of “US” and “AU”. This could assist us in case of changing our strategy to focus on these two countries for more commercial approaches by region.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, ορθογώνιο παραλληλόγραμμο

Περιγραφή που δημιουργήθηκε αυτόματα

Second, we analyzed the performance of the products. The visualization of sales per product showed that some products have much more sales than others, this information could be utilized by the directors for commercial and marketing purposes which will either change the sales strategy for the less sold products or use the more sold products as an assistance for the less sold products.

Εικόνα που περιέχει διάγραμμα, στιγμιότυπο οθόνης, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

Third, we examined the distribution of the children per household and how this affects the sales per customer is a significant aspect of the analysis as it provides insights into the composition of households. The number of children in a household can impact the needs of products and services which will assist us in shaping strategies for our target group.

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, ορθογώνιο παραλληλόγραμμο

Περιγραφή που δημιουργήθηκε αυτόματα

Furthermore, we took the distribution of the customers according to their yearly income which will provide us the insight into the economic structure of our customers. It is a crucial step in understanding how income affects purchasing behavior. This analysis can lead to tailored and personalized sales, offers, approaches, and services.

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Fourth, we analyzed the geographic distribution, product performance, sales prediction, and customer segmentation will provide us with a comprehensive framework for decision-making. It is crucial to pay attention to the distributions of children and customers based on income, hence they offer strategic opportunities for adaptation and growth.

Εικόνα που περιέχει κείμενο, κερί, πολυχρωμία, μολύβι

Περιγραφή που δημιουργήθηκε αυτόματα

Fifth, we examined the distribution of annual incomes per city which reveals significant aspects of the economic reality in different regions. Through this analysis, potential income gaps between cities can be identified and we can understand the economic pattern of each area. According to these insights via the visual representation, a business can adopt strategies to address customer needs in various financial conditions. Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Moreover, the analysis of sales amount will provide us with the tendency to the purchases between the two genders. We used this visualization for the purpose of understanding better which products and services are more popular between the two genders and to customize our approach by getting the assistance of the marketing sector.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, διάγραμμα, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Finally, I used linear regression for the purpose of estimating the relationship between a dependent variable and one or more independent variables. It is a statistical technique which we used to assess the characteristics of ‘Yearly Income’ , ‘Total Product cost’, ‘Annual Sales’ and ‘Unit Price’ relate to the dependent variable ‘Sales Amount’. The steps we followed was the data preparation, model training and evaluation, interpretation of coefficients, adjustment and optimization. It is necessary to understand the economic patterns which assists to personalize approaches based on these patterns of each area.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμμή, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμμή, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματαΕικόνα που περιέχει κείμενο, διάγραμμα, γράφημα, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα

The previous analysis shows the importance of understanding the demographic and economic aspects of our customers. The main picture results from this analysis can enhance decision-making capabilities, giving the opportunity to businesses to adapt their strategies for optimizing performance and responsiveness to market needs.

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, γράφημα

Περιγραφή που δημιουργήθηκε αυτόματα

Conclusion

To Conclude, I followed all the steps to analyze a dataset in case to evaluate these insights, gain important information from these, and achieve the goal of processing this big dataset by using MySQL and Python in order to proceed with accurate findings. In this assignment I faced challenges, it was the first time that I combined Python and MySQL for such big data, and it was a challenge to provide accurate visualizations with clear insights. It is an assignment that needed a lot of documentation about Python and MySQL and of course machine learning which I used for the first time in an assessment in a huge dataset and it was a great experience and joyful.

In my opinion, it was a great lesson that maybe needs more time to delve deeply into Python, cloud computing, and libraries that we use for processing big data.