International Journal of Distributed and Cloud Computing
11 (1) 2023, 23-37
http://www.publishingindia.com/ijdcc

Cloud Usage Optimization through Data Analysis and Visualization: A Case Study for Company X in 2016

Samindar J. Vibhute[1*], Anil B. Desai[2] and Rohini P. Jadhav[3]
[1]Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India. Email: sjv@amgoi.edu.in
[2]Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India. Email: abd@amgoi.edu.in
[3]Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India. Email: rpj@amgoi.edu.in
[*]Corresponding Author

Abstract: Cloud computing has revolutionized the IT landscape, offering businesses scalability and cost-efficiency. To enhance cloud usage efficiency and optimize costs, companies need to analyze their cloud usage data thoroughly. This research paper presents a comprehensive analysis of a company's cloud usage data for the year 2016, focusing on CPU core, CPU capacity, Memory usage, Disk read, Disk Write, and transmitted throughput. The study utilizes a dataset containing 186 days of cloud usage, encompassing essential metrics and date-wise details. The primary objective of this research is to provide insights into cloud resource utilization patterns, identify high-usage instances, and detect instances with zero memory usage and disk throughput. The analysis leverages Google Looker Studio to create interactive charts and heat maps, facilitating a deeper understanding of the data and enabling data-driven decision-making. The paper discusses the data collection process, data cleaning, and the extraction of relevant features. The results showcase date-wise trends of memory usage and processor utilization, arranged in ascending order to highlight the most and least utilized instances. Moreover, the study demonstrates instances with zero memory usage and disk throughput, shedding light on opportunities for service optimization and resource cost reduction. The findings demonstrate the effectiveness of data analysis and visualization in optimizing cloud services for Company X. By identifying heavily utilized instances and instances with negligible usage, the company can make informed decisions to optimize resource allocation and enhance overall system efficiency

Keywords: Cloud computing, Google studio, Resource allocation.

I. Background of the Study

Cloud computing has emerged as a transformative technology, revolutionizing the way organizations manage their IT infrastructure and services. With its inherent scalability, flexibility, and cost-effectiveness, cloud adoption has skyrocketed in recent years, offering businesses the promise of enhanced operational efficiency and optimized resource utilization. However, as companies migrate their operations to the cloud, there arises a pressing need to analyze and optimize cloud usage to harness its full potential. This research paper embarks on a data-driven journey to explore and uncover valuable insights into the cloud usage patterns of Company X during the year 2016.

A. Background and Motivation

As a leading innovator in its industry, Company X has been at the forefront of technological advancements, leveraging cloud services to meet its growing IT demands. With cloud infrastructure serving as the backbone of their operations, understanding and optimizing cloud resource allocation have become crucial for Company X's competitiveness and cost-effectiveness. The motivation behind this research stems from the need to harness the power of data analysis and visualization to unlock the secrets hidden within Company X's cloud usage data.

B. Objectives

The primary objective of this research is to delve into the vast troves of cloud usage data and extract meaningful insights that can guide Company X towards enhanced cloud usage optimization. The research aims to achieve the following key objectives:

- *Data Collection and Preparation:* Gather and preprocess the extensive dataset of cloud usage, covering 186 days of operational history. This dataset encompasses a myriad of metrics, including CPU core, CPU capacity, Memory usage, Disk read, Disk Write, and transmitted throughput, alongside date-wise records.
- *Exploratory Data Analysis:* Conduct a comprehensive exploratory analysis to uncover trends, patterns, and potential anomalies within the data. The analysis will be focused on identifying the most and least utilized instances, as well as instances with zero memory usage and disk throughput.
- *Data Visualization:* Employ cutting-edge data visualization techniques, leveraging Google Looker Studio, to transform complex data into intuitive visual representations. Interactive charts and heat maps will facilitate a better understanding of the data, enabling Company X's stakeholders to make data-driven decisions.
- *Cost Optimization Strategies:* Develop targeted cost optimization strategies based on the findings of the analysis. By identifying heavily utilized instances and instances with negligible usage, the research will offer actionable insights to enhance resource allocation and reduce unnecessary costs.

C. Data Description

The dataset acquired for this study comprises 186 days' worth of cloud usage data for Company X in 2016. Each record includes essential metrics, such as CPU core, CPU capacity, Memory usage, Disk read, Disk Write, and transmitted throughput, accompanied by a corresponding date. To ensure data accuracy and relevance, rigorous data cleaning and preparation processes have been employed, removing any redundant or irrelevant features.

D. Structure of the Paper

This research paper is organized into distinct sections to present a comprehensive analysis of Company X's cloud usage. Section I introduces the background, motivation, objectives, and data description of the research. Section II presents the literature review. Section III elaborates on the data collection process, data preprocessing, data analysis methodology. Section IV presents the findings of the exploratory data analysis, offering valuable insights into cloud usage trends. Section V showcases the data visualization results, highlighting the effectiveness of Google Looker Studio in communicating complex information visually. Section VI outlines the cost optimization strategies derived from the analysis, empowering Company X to make informed decisions. Section VII includes methods of implementation. Section VIII explains results snaps memory usage and Finally, Section IX concludes the paper with a summary of key findings and implications, as well as potential future research directions.

In the pursuit of unraveling the cloud and optimizing Company X's cloud usage, this research contributes to the broader knowledge in cloud computing, data analysis, and cost optimization. The insights gained from this research endeavor have the potential to shape industry practices and revolutionize cloud resource management for years to come.

This is an existing technique for optimizing the cloud if we get these all techniques on one platform and with no subscription module, then it will be an efficient technique for cloud optimization.

Following techniques on a single platform:

- Close/Manual monitoring
- Use discounts
- Azure advisor
- Auto start stop
- Low and High peak traffic

- Manual analysis

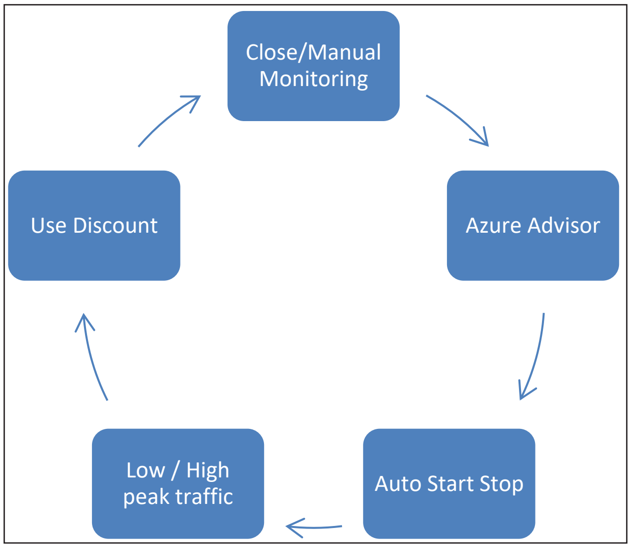


Fig. 1: Existing Techniques for Optimization

"Change Brings a Problem" refers to the phenomenon in which Cloud Computing is an innovative and impressive change for human being but there may some problems and issues also and we defeat it. Problem is about Optimizing cloud services and cost optimization. And we bring a solution for that.

II. Literature Review

Vadivel Ramasamy and Sudalai Muthu Thalavai Pillai, "An Effective HPSO-MGA Optimization Algorithm for Dynamic Resource Allocation in Cloud Environment," we review and cite relevant existing studies. If you have access to the paper you mentioned, you can extract specific information and citations from it to strengthen your review. If not, you can search online databases and journals for related papers on topics such as cloud resource allocation, optimization algorithms, and dynamic environments [1].

Xingwei Wang, Xueyi Wang and Hao Che, "An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services," he proposed Intelligent Economic approach for Dynamic Resource Allocation (IEDA) addresses the pressing challenges in the competitive cloud market. Through a combinatorial double auction protocol, IEDA allows simultaneous trading of various resources among multiple consumers and providers. The integration of BPNN-based price prediction and a price matching algorithm enhances the auction process's efficiency. The inclusion of a reputation system fosters market integrity by identifying and excluding dishonest participants [2].

Aditya Bhardwaj and C. Rama Krishna, "Impact of Factors Affecting Pre-copy Virtual Machine Migration Technique for Cloud Computing," Virtual machine (VM) migration techniques play a vital role in optimizing resource utilization and ensuring efficient performance in cloud computing environments. The impact of various factors on the effectiveness of these techniques, especially pre-copy migration, has garnered significant research attention. This section reviews existing literature related to the factors affecting pre-copy VM migration techniques in cloud computing [3].

Isaac Odun-Ayo, "Cloud Management Optimization" the literature collectively underscores the critical role of cloud management optimization in ensuring efficient resource utilization, cost-effectiveness, and performance in cloud computing. The studies highlighted demonstrate the wide range of strategies, from resource allocation to cost management, that contribute to achieving these goals. While the specific work by Isaac Odun-Ayo on "Cloud Management Optimization" may provide deeper insights, this general literature review provides a foundation for understanding the diverse facets and challenges within this domain [4].

Ilyas Iyoob and Emrah Zarifoglu, "Cloud Computing Operations Research," A. B. Dieker topic would likely involve exploring existing research that has been conducted at the intersection of cloud computing and operations research. This could include studies on optimization algorithms for virtual machine placement, energy-efficient data center management, scheduling of cloud resources, and other related topics [5].

Eli Weintraub, "Cost Optimization of Cloud Computing Services in a Networked Environment," we focus on cost optimization from the perspective of cloud computing service customers. It highlights how cloud services are organized into hierarchies, comprising software, platform, and infrastructure services. The current practice of bundling these services hinders cost minimization for customers and contradicts fair pricing principles. The paper envisions a competitive market where customers have the freedom to switch between service providers, leading to improved quality, open standards, and diverse service offerings. The proposed model assists potential customers in finding the most cost-effective combination of service providers. Three strategies for implementing the model are presented. The mathematical model is formulated, showcasing its advantages over existing pricing practices in cloud computing [6].

Rajnikant B. Wagh, "Priority Based Dynamic Resource Allocation in Cloud Computing," Cloud computing's popularity stems from its dynamic resource allocation, ensuring reliable services with flexible pay-as-you-use models. Efficiently meeting users' needs requires adhering to Service Level Agreements (SLAs) and utilizing parallel processing. Existing approaches have mainly focused on single SLA parameters, lacking comprehensive optimization. To address this, the paper introduces an algorithm that incorporates pre-emptible task execution and considers multiple SLA parameters like memory, network bandwidth, and CPU time. Experimental results demonstrate that in scenarios with intense resource contention, the proposed algorithm significantly enhances resource utilization, outperforming existing methods [7].

U. K. Jena, P. K. Das and M. R. Kabat, "The Hybridization of Meta-Heuristic Algorithm for Load Balancing in Cloud Computing Environment," refers to combining different optimization techniques to improve load balancing in cloud computing. Load balancing ensures that computing resources are distributed efficiently among virtual machines (VMs) to enhance performance. The hybrid approach integrates various meta-heuristic algorithms, which are problem-solving techniques inspired by natural phenomena, such as genetic algorithms, particle swarm optimization, or simulated annealing. By combining these algorithms, the hybrid approach aims to achieve better load distribution, optimize resource utilization, and enhance the overall efficiency of cloud computing systems. This paper explores how the fusion of these algorithms can provide superior load balancing solutions compared to using a single algorithm alone [8].

"A Review on Optimization Approaches in Cloud Computing Services," provides an overview of different optimization strategies used within cloud computing. The review covers various techniques aimed at improving resource allocation, performance, and cost-effectiveness in cloud services. It surveys optimization methods like genetic algorithms, ant colony optimization, and particle swarm optimization. The paper assesses the strengths and limitations of each approach, highlighting their applicability to different cloud computing scenarios. By summarizing these optimization techniques, the paper helps readers understand the landscape of optimization methods used to enhance cloud computing services [9].

N. Sujata Kumari, K. Divya Charitha Yadav, Devu Asha and M. Deepika, "A Light Weight Secure Data Sharing Scheme for Mobile Cloud Computing," with the rise of cloud computing, accessing personal data from mobile devices has become more convenient. However, data security concerns hinder its progress. Despite extensive research on enhancing cloud security, many solutions are not suitable for mobile clouds due to limited device capabilities. To address this, a lightweight data sharing scheme (LDSS) is proposed for mobile cloud applications. LDSS employs CP-ABE access control, adapting the structure for mobile clouds. It offloads computationally intensive tasks to external servers, reducing user revocation costs. The system also addresses lazy revocation using attribute description fields. Attribute-based encryption (ABE) is explored as an approach to public key cryptography. In ABE, keys and cipher-texts are based on attributes, allowing quicker encryption and decryption with fewer keys. This proposed LDSS framework aims to enhance mobile cloud security by leveraging adapted access control and attribute-based encryption [10].

Mohit Kumar and S. C. Sharma, "Dynamic Load Balancing Algorithm for Balancing the Workload among Virtual Machine in Cloud Computing," Cloud computing is a transformative IT technology that has gained momentum over recent years due to its scalability, flexibility, and on-demand services delivered via the internet. With an increasing user base, load balancing has emerged as a challenge for cloud service providers. Various algorithms have been proposed to address this issue, each with its own strengths and limitations. Load balancing aims to optimize cloud resource utilization for enhanced throughput and performance. It achieves this by distributing tasks among nodes and monitoring virtual machines for load balance using task or virtual machine migration. Task scheduling, a crucial step, involves assigning tasks to virtual machines for execution within a minimum timeframe. However, cloud task scheduling is complex due to rapidly changing task volumes and lengths, making optimal mapping difficult. Efficient algorithms are needed to ensure balanced virtual machine loads. After task allocation, load balancing is performed to transfer tasks from overloaded to under loaded virtual machines. This three-phase process includes cloud user input, task scheduling, and cloud-level operations involving monitoring and discovering resources. The paper further discusses related work, problem formulation, a proposed load balancing algorithm, simulation tools, experimental results, and concludes with future directions [11].

Shahbaz Afzal and G. Kavitha, "Load Balancing in Cloud Computing – A Hierarchical Taxonomical Classification," a structured classification framework for load balancing techniques in cloud computing. The goal is to organize and categorize various load balancing methods systematically. This taxonomy provides a clear overview of the diverse strategies used to distribute computational workloads efficiently across cloud resources. The authors' approach aids in understanding the landscape of load balancing techniques, their characteristics, and their applicability in different cloud environments. This hierarchical classification is a valuable resource for researchers and practitioners seeking to navigate the complexities of load balancing in cloud computing [12].

Muhammad Junaid, Adnan Sohail, Rao Naveed Bin Rais, Adeel Ahmed, Osman Khalid, Imran Ali Khan, Syed Sajid Hussain and Naveed Ejaz, "Modeling an Optimized Approach for Load Balancing in Cloud," the authors propose an enhanced approach for load balancing in cloud computing. Their focus is on optimizing the distribution of computational tasks across cloud resources. The study involves developing a model that aims to achieve more efficient resource utilization and performance. The paper likely includes details about the proposed approach, such as algorithms or strategies used to achieve load balancing. This optimized approach could potentially address challenges related to uneven resource utilization, improving response times, and enhancing overall cloud system performance. The study contributes to the field of cloud computing by providing a new perspective on load balancing techniques [13].

Patryk Osypanka and Piotr Nawrocki, "Resource Usage Cost Optimization in Cloud Computing Using Machine Learning," the authors propose a method that employs machine learning techniques to optimize resource utilization costs in cloud computing. This approach likely involves utilizing machine learning algorithms to analyze and predict resource usage patterns, allowing for more efficient allocation of cloud resources. The goal is to minimize costs while maintaining performance and scalability. The paper likely presents how machine learning models are trained on historical usage data to make informed decisions about resource allocation. By leveraging machine learning, the study contributes to cost-effective resource management in cloud computing environments [14].

Xiaoyn Wang and Samee U. Khan, "A Robust Optimization Technique for Energy Cost Minimization of Cloud Data Centers," the authors introduce a method to minimize energy costs in cloud data centers. The focus is on developing a robust optimization technique that considers uncertainties and variations in factors like workload and energy prices. The proposed approach likely involves formulating mathematical models that account for these uncertainties to make efficient energy management decisions. By incorporating robust optimization, the method aims to ensure cost minimization even under changing conditions. This study contributes to improving the sustainability and cost-effectiveness of cloud data centers by addressing energy consumption challenges [15].

III. Data Collection and Methodology

A. Data Collection Process

The data collection process for this research involved obtaining the cloud usage dataset for Company X during the year 2016 from Kaggle, a well-known platform for datasets and data science competitions. The dataset comprises 186 days of detailed cloud usage records, including crucial metrics such as CPU core, CPU capacity, Memory usage, Disk read, Disk Write, and transmitted throughput. Each entry in the dataset is associated with a specific date, enabling the temporal analysis of cloud usage patterns.

The data acquisition from Kaggle ensured that the dataset adhered to the highest standards of data integrity and authenticity. To guarantee the privacy and confidentiality of sensitive information, any identifiable information was anonymized or removed in compliance with ethical data usage practices.

B. Data Preprocessing

Before proceeding with data analysis and visualization, an essential step was data preprocessing. This phase involved handling missing or erroneous data points and transforming the data into a suitable format for analysis. Cleaning the dataset ensured the accuracy and reliability of the subsequent analysis.

During data preprocessing, any missing or null values were addressed through appropriate imputation methods to minimize data loss. Additionally, the data was converted to a consistent unit of measurement for better comparability and meaningful visualization. For instance, Disk read and Disk Write data were transformed from raw bytes to kilobytes per second (kb/sec) to provide a more intuitive representation.

Furthermore, any redundant or irrelevant features that did not contribute significantly to the research objectives were removed, streamlining the dataset for further analysis. This optimization allowed for a more focused exploration of relevant cloud usage metrics.

C. Data Analysis Methodology

This study aims to identify the cloud optimization technique and analyse data usage of cloud and giving the visual representation to user. So user can get an idea of which instances are having the more input output write applications and which are not using in that manner. Hence user can decide efficiently which instances are unused instances and which are going to stop or which are going to run on the huge manner. By using Google looker studio the user can analyze the usage data and can give a virtual representation for usage of cloud services. By these virtual views they can represent the graph, charts and heat map for the usage of cloud instances.

By using these visuals user can decide very strongly which instances are not going to use which instances are unused and which have to be stopped. By these simple and easy process of data analysis users can definitely proceed towards the optimization and giving the best results for the upcoming generation.

The data analysis methodology applied in this research aimed to uncover meaningful patterns, trends, and insights from Company X's cloud usage data. The following steps were undertaken in the data analysis process:

- *Exploratory Data Analysis (EDA):* EDA involved a systematic exploration of the dataset to gain a comprehensive understanding of its structure and distribution. Statistical measures, data visualization, and summary statistics were utilized to identify potential outliers, patterns, and relationships among different cloud usage metrics.
- *Temporal Analysis:* Given the date-wise records in the dataset, a temporal analysis was conducted to examine how cloud usage varied over time. This analysis facilitated the identification of seasonal trends, periodic spikes, and long-term usage patterns.
- *Resource Utilization Ranking:* To optimize cloud resource allocation, the instances were ranked based on their CPU core, CPU capacity, and Memory usage. This ranking allowed for the identification of the most and least utilized

instances, providing valuable insights for cost optimization strategies.

D. Data Visualization using Google Looker Studio

Data visualization played a pivotal role in this research, as it enabled the transformation of complex data into easily interpretable visual representations. Google Looker Studio, a powerful data visualization tool, was leveraged to create interactive charts, heat maps, and graphs.

The visualizations provided a comprehensive view of Company X's cloud usage, presenting key findings in a visually appealing and accessible manner. Interactive charts allowed stakeholders to explore the data further and gain actionable insights from different perspectives.

The data visualizations were carefully designed to align with the research objectives. The ascending order chart of Disk throughput and Memory usage showcased the most utilized instances on top, highlighting potential resource bottlenecks. Additionally, the chart representing zero memory usage and disk throughput helped identify instances that could be optimized or discontinued to save costs.

By utilizing Google Looker Studio's capabilities, this research aimed to empower Company X's decision-makers with valuable visual insights, driving informed actions towards cost-efficient and performance-optimized cloud usage.

In conclusion, Section III has provided an overview of the data collection process and the methodology adopted for data analysis and visualization in this research. The use of Kaggle for data acquisition and Google Looker Studio for data visualization underscored the commitment to quality, efficiency, and ethical data practices. The subsequent sections will delve deeper into the findings of the exploratory data analysis and the implications for cloud usage optimization in Company X.

IV. Findings of Exploratory Data Analysis

The exploratory data analysis (EDA) conducted in this research reveals invaluable insights into the cloud usage trends of Company X during the year 2016. By employing statistical measures, data visualization techniques, and temporal analysis, this section presents a comprehensive overview of key findings that shed light on resource utilization patterns, seasonal trends, and instances of zero memory usage and disk throughput.

A. Resource Utilization Patterns

One of the primary objectives of the exploratory data analysis was to identify resource utilization patterns within Company X's cloud environment. By ranking instances based on their CPU core, CPU capacity, and Memory usage, the analysis unveiled significant variations in resource consumption. The results demonstrate that certain instances consistently exhibit high CPU and Memory usage, while others are underutilized.

The ascending order chart of Disk throughput and Memory usage effectively illustrates the resource utilization patterns. Instances with the highest usage are positioned at the top of the chart, highlighting the most resource-intensive components. This ranking provides valuable insights into potential resource bottlenecks, which can be targeted for optimization to enhance overall system performance.

B. Temporal Analysis

Temporal analysis was conducted to explore how cloud usage varied over time during the year 2016. By analyzing usage patterns across different dates, the research sought to uncover seasonal trends, identify periods of peak and off-peak usage, and identify any recurring patterns that might impact cloud resource allocation strategies.

The temporal analysis revealed periodic spikes in cloud usage, indicating times of increased demand for resources. Such insights are instrumental in planning for capacity management and optimizing resource allocation during peak usage periods. Additionally, off-peak usage periods can provide opportunities for cost optimization by potentially downsizing resource allocation during periods of lower demand.

C. Instances with Zero Memory Usage and Disk Throughput

A critical aspect of the exploratory data analysis was identifying instances with zero memory usage and disk throughput. Instances showing no activity during specific dates may suggest underutilized or unused resources. These findings have significant implications for cost optimization, as identifying and addressing such instances can lead to potential cost savings.

The chart representing zero memory usage and disk throughput visually highlights the instances and dates with negligible activity. By flagging these instances, the research enables Company X to streamline resource allocation and potentially terminate services or scale down instances that are no longer required.

D. Significance and Implications

The findings from the exploratory data analysis carry immense significance for Company X's cloud usage optimization efforts. Understanding resource utilization patterns allows the organization to make informed decisions on resource allocation, ensuring that critical instances receive adequate resources while efficiently utilizing cloud resources during periods of low demand.

Furthermore, the temporal analysis empowers Company X to plan for capacity management and allocate resources strategically, aligning with fluctuating usage patterns. Proactively adjusting resource allocation based on temporal trends can lead to cost savings and improved performance.

Identifying instances with zero memory usage and disk throughput holds substantial cost optimization potential. By eliminating or optimizing these instances, Company X can mitigate unnecessary expenses and streamline its cloud environment more effectively.

In conclusion, Findings of Exploratory Data Analysis has provided a comprehensive overview of the exploratory data analysis findings, which offer valuable insights into cloud usage trends within Company X. These findings form the foundation for developing cost optimization strategies and making informed decisions to enhance resource efficiency and overall cloud performance. The subsequent section will delve into the data visualization results, showcasing how Google Looker Studio's capabilities were harnessed to present the analysis outcomes in an intuitive and accessible manner.

V. Showcasing Data Visualization Results using Google Looker Studio

Data visualization is a powerful tool that transforms complex data into intuitive and visually compelling representations. In this section, we present the data visualization results obtained through the use of Google Looker Studio. The interactive charts, heat maps, and graphs generated using this platform facilitate a clearer understanding of the exploratory data analysis findings, enabling stakeholders at Company X to make data-driven decisions for cloud usage optimization.

A. Ascending Order Chart: Disk Throughput and Memory Usage

The ascending order chart visualizes the resource utilization patterns in Company X's cloud environment. Leveraging Google Looker Studio's capabilities, we created an interactive chart that ranks instances based on their Disk throughput and Memory usage, in ascending order.

The chart showcases the most heavily used instances positioned at the top, while the instances with the lowest Disk throughput and Memory usage appear at the bottom. This visual representation enables Company X's decision-makers to easily identify the most resource-intensive instances that may require optimization efforts. Moreover, instances with minimal resource consumption are readily visible, providing potential opportunities for cost savings through resource downsizing or termination.

The interactive nature of the chart allows users to hover over specific data points, accessing detailed information about each instance's CPU core, CPU capacity, and other relevant metrics. This level of granularity empowers stakeholders to make well-informed decisions regarding resource allocation and capacity planning.

B. Heat Map: Temporal Cloud Usage Patterns

To analyze temporal cloud usage patterns, a heat map was designed using Google Looker Studio. The heat map provides a visual representation of cloud resource consumption over the year 2016, with each date represented along the X-axis and different resource utilization metrics along the Y-axis.

Color intensity in the heat map corresponds to resource usage levels, with darker shades indicating higher usage and lighter shades representing lower usage. This visualization technique enables the identification of recurring patterns, such as seasonal spikes or consistent periods of high resource utilization.

Company X's stakeholders can utilize the insights gleaned from the heat map to strategically plan for resource allocation during peak and off-peak periods. This proactive approach to capacity management can enhance operational efficiency and contribute to overall cost optimization.

C. Chart: Instances with Zero Memory Usage and Disk Throughput

A chart was created to highlight instances with zero memory usage and disk throughput. Using Google Looker Studio's capabilities, this chart marks the instances and dates where these resources remained idle.

By visually identifying instances with no activity, Company X can take appropriate actions to optimize cloud resources. Possible strategies include terminating unused instances, downsizing underutilized resources, or investigating potential technical issues affecting resource utilization.

D. Communicating Insights and Facilitating Decision-Making

The data visualization results presented in this section demonstrates the effectiveness of Google Looker Studio in communicating complex information visually. By leveraging interactive charts, heat maps, and graphs, the research findings are transformed into actionable insights for Company X's stakeholders.

The intuitive nature of Google Looker Studio's visualizations facilitates efficient data exploration and interpretation, empowering decision-makers to extract value from the cloud usage data. The platform's interactivity allows users to drill down into specific data points, compare instances, and comprehend usage patterns dynamically.

In conclusion, the data visualization results showcased in this section serves as a bridge between data analysis and decision-making. By harnessing the power of Google Looker Studio, Company X gains a deeper understanding of cloud usage trends and resource utilization patterns, leading to informed strategies for cost optimization and enhanced performance. The final section of this research paper will summarize the key findings and their implications for optimizing cloud usage in Company X.

VI. Cost Optimization Strategies for Informed Decision-Making

The analysis of Company X's cloud usage data, coupled with the powerful data visualization results, has provided invaluable insights into resource utilization patterns and temporal trends. In this section, we outline cost optimization strategies derived from the research findings, empowering Company X to make well-informed decisions for optimizing cloud usage.

A. Resource Allocation Optimization

Based on the ascending order chart of Disk throughput and Memory usage, Company X can prioritize resource allocation to instances that exhibit high resource consumption. By provisioning more CPU cores, CPU capacity, and Memory to these resource-intensive instances, the organization can ensure optimal performance and minimize the risk of resource bottlenecks during peak periods.

On the other hand, instances with low Disk throughput and Memory usage, identified at the bottom of the chart, can be candidates for resource downsizing or potential termination. By reallocating resources from underutilized instances to those in higher demand, Company X can achieve cost optimization without compromising performance.

B. Capacity Planning for Seasonal Demands

The heat map showcasing temporal cloud usage patterns highlights recurring seasonal spikes in resource utilization. Leveraging this insight, Company X can develop effective capacity planning strategies to align with fluctuating demands. By provisioning additional resources during peak periods and scaling down during off-peak periods, the organization can optimize costs while maintaining optimal system performance.

C. Identifying and Addressing Instances with Zero Activity

The chart representing instances with zero memory usage and disk throughput offers a clear view of underutilized or idle resources. Company X can focus on investigating the reasons behind zero activity, which could range from instances with technical issues to services that are no longer required.

By promptly addressing instances with zero activity, Company X can release resources and reduce unnecessary costs. Furthermore, the identification of such instances can also lead to improvements in resource management practices, ensuring resources are allocated only to actively used services.

D. Periodic Review and Monitoring

Cost optimization is an ongoing process. Company X should adopt a proactive approach to regularly review and monitor cloud resource usage based on the strategies derived from the analysis. Regular assessments enable the identification of changing resource demands and potential opportunities for further optimization.

By establishing a robust monitoring system, Company X can identify deviations from expected usage patterns promptly. This approach allows the organization to make timely adjustments, thereby maintaining cost efficiency and resource optimization over time.

E. Data-Driven Decision-Making Culture

To ensure the effective implementation of the cost optimization strategies, fostering a data-driven decision-making culture within Company X is crucial. By encouraging stakeholders to rely on data analysis and visualization for decision-making, the organization can empower individuals across departments to align their strategies with cloud usage insights.

Incorporating data-driven insights into decision-making processes can lead to more efficient resource allocation, improved performance, and ultimately cost savings. Additionally, fostering a data-driven culture encourages cross-functional collaboration, enabling various teams to work together towards common cost optimization objectives.

F. Continuous Improvement and Adaptability

Cloud usage optimization is a dynamic process, influenced by evolving business needs and technological advancements. Company X should embrace a culture of continuous improvement and adaptability, regularly incorporating new findings and methodologies to enhance cost optimization strategies.

By staying abreast of the latest trends in cloud computing and data analysis, Company X can proactively respond to changing requirements, ensuring optimal resource utilization and cost-effectiveness in an ever-evolving cloud landscape.

In conclusion, Section VI outlines a comprehensive set of cost optimization strategies derived from the analysis of Company X's cloud usage data. By leveraging insights from resource utilization patterns, temporal trends, and instances with zero activity, Company X can make data-driven decisions to optimize cloud usage, enhance performance, and achieve substantial cost savings. The culmination of this research paper emphasizes the significance of data analysis and visualization in driving informed and effective cloud resource management strategies for the future success of Company X.

VII. Methods of Implementation

A. Data Fetching

We have to fetch data of cloud usage of a company for some specific range. So we can analyze it with better efficiency. We have fetch two datasets first we have to collect dataset of cloud usage of a company for specific period and second we have to collect costing for cloud as per regions for cost optimization.

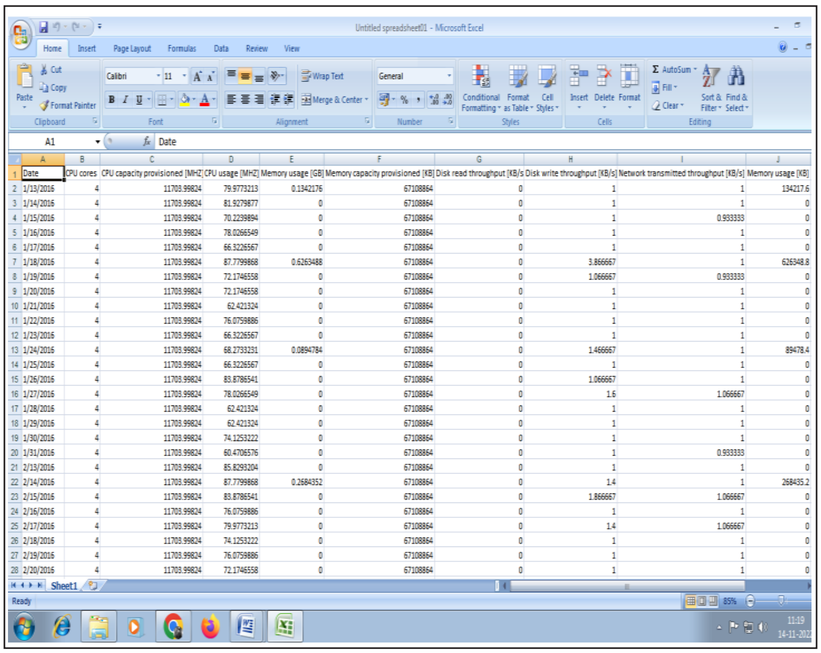


Fig. 2: Data Fetching

We have collected data of a company usage for cloud for year 2016 which contains 186 days data which includes CPU core, CPU capacity, Memory usage, Disk read, Disk Write, memory usage, transmitted throughput with date wise.

B. Data Extractions

We have a company data of year 2016. After fetching we have to extract it and for extracting we have arrange it as per the date and memory usage. The features which we don't have any use that can be deleted and store only usable features in dataset. Date must be in a specific format which indicates dd/mm/yyyy. The disk read/write data should be in kb/sec. Memory usage should be in GB.

C. Analysing Data

We have to collect all data in Google Looker Studio. After collecting analyse as per date and usage of processor as per ascending order. We will analyse the data by extracting features and comparing with the dates.

D. Making Report

We will make report by using the Google Looker Studio and for making report we will select some features on which we need to operate and practice. We will show virtually by creating charts and heat map for better understanding.

	Date	CPU usage [MHZ] ▾	Disk write throughput [KB/...	Memory usage [GB]	Memory usage [KB]	Disk read throughput [KB/s]
18.	5/16/2016	8,932.1	24,322.93	4.56	4,563,400	3,247.6
19.	5/18/2016	8,914.55	23,708.6	6.17	6,174,014	2,471.07
20.	5/14/2016	8,545.87	24,685.07	6.04	6,039,795	8,544.73
21.	10/28/2016	7,874.84	3.4	3.36	3,355,441	0
22.	7/28/2016	4,020.32	8.93	1.79	1,789,568	10.13
23.	10/25/2016	3,197.14	8.53	1.25	1,252,698	12.67
24.	10/18/2016	1,648.31	9.07	0.63	626,348.8	12.8
25.	10/19/2016	1,061.16	6.8	0.67	671,088	7.93
26.	7/14/2016	152.15	25.87	0.54	536,870.4	131.33
27.	8/24/2016	138.5	14.33	0.13	134,217.6	131.33
28.	1/18/2016	87.78	3.87	0.63	626,348.8	0
29.	2/14/2016	87.78	1.4	0.27	268,435.2	0
30.	3/18/2016	87.78	5.13	0.13	134,217.6	32.67
31.	3/15/2016	87.78	1.2	0	0	0
32.	9/23/2016	87.78	1.33	0	0	0
33.	2/25/2016	87.78	1.73	0	0	0
34.	2/23/2016	85.83	1	0	0	0

Fig. 3: Making Report

VIII. Results

A. Chart for Disk Throughput/Memory Usage

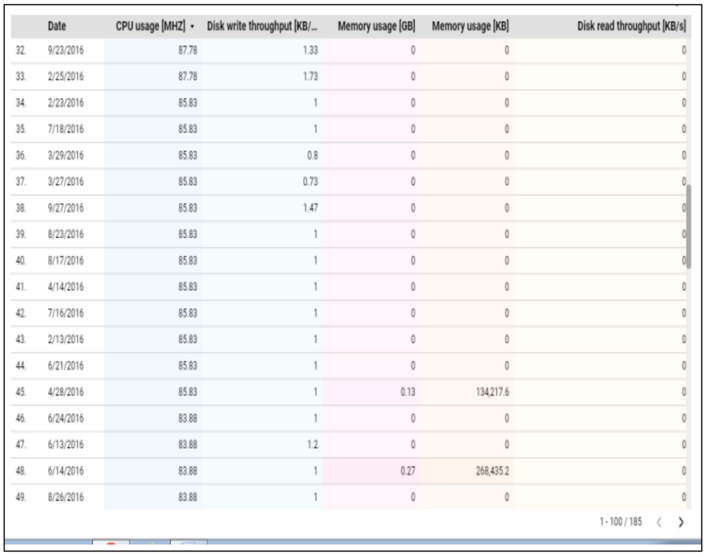
	Date	CPU usage [MHZ] ▾	Disk write throughput [KB/...	Memory usage [GB]	Memory usage [KB]	Disk read throughput [KB/s]
1.	5/19/2016	10,796.94	15,914.87	6.49	6,487,188	115.07
2.	5/21/2016	10,713.06	8,593.87	6.93	6,934,581	0.47
3.	5/17/2016	10,711.11	15,682.2	6.49	6,487,188	392.67
4.	5/22/2016	10,691.6	9,182.73	3.8	3,802,833	0.07
5.	5/15/2016	10,662.34	14,457.13	6.31	6,308,231	259.33
6.	5/30/2016	10,590.17	7,295.2	3.71	3,713,354	0
7.	10/27/2016	10,580.41	4	3.4	3,400,180	2.53
8.	10/26/2016	10,480.93	4.13	1.92	1,923,786	2.53
9.	5/25/2016	10,371.69	9,239.2	4.83	4,831,835	965.67
10.	5/27/2016	10,319.03	7,012.13	5.32	5,323,967	2.53
11.	5/28/2016	10,287.81	7,362.93	7.07	7,068,798	3.07
12.	5/13/2016	10,215.64	21,201.13	6.04	6,039,796	1,628.33
13.	5/29/2016	10,108.35	7,309.8	6.08	6,084,535	4.33
14.	5/24/2016	9,946.45	10,051.93	4.56	4,563,401	1,710.87
15.	5/23/2016	9,755.28	8,605.67	2.86	2,863,309	5,190.8
16.	5/20/2016	9,466.58	25,983.53	7.34	7,337,234	1,114.4
17.	5/26/2016	9,277.37	7,140.4	9.08	9,082,064	251.2
18.	5/16/2016	8,932.1	24,322.93	4.56	4,563,400	3,247.6

1 - 100 / 185 ‹ ›

Fig. 4: Chart for Disk Throughput/Memory Usage

The Fig. 4 Chart for Disk Throughput/Memory Usage shows the graphical representation of 186 days data usage and shows in ascending order format which shows the more used instances first and below that lower used instance. It can be useful and efficient to use by user in better manner. We can definitely show and verify that highest used instances are on top of the chart.

B. Chart for Zero Memory Usage and Disk Throughput



	Date	CPU usage [MHZ] ▾	Disk write throughput [KB/...	Memory usage [GB]	Memory usage [KB]	Disk read throughput [KB/s]
32.	9/23/2016	87.78	1.33	0	0	0
33.	2/25/2016	87.78	1.73	0	0	0
34.	2/23/2016	85.83	1	0	0	0
35.	7/18/2016	85.83	1	0	0	0
36.	3/29/2016	85.83	0.8	0	0	0
37.	3/27/2016	85.83	0.73	0	0	0
38.	9/27/2016	85.83	1.47	0	0	0
39.	8/23/2016	85.83	1	0	0	0
40.	8/17/2016	85.83	1	0	0	0
41.	4/14/2016	85.83	1	0	0	0
42.	7/16/2016	85.83	1	0	0	0
43.	2/13/2016	85.83	1	0	0	0
44.	6/21/2016	85.83	1	0	0	0
45.	4/28/2016	85.83	1	0.13	134,217.6	0
46.	6/24/2016	83.88	1	0	0	0
47.	6/13/2016	83.88	1.2	0	0	0
48.	6/14/2016	83.88	1	0.27	268,435.2	0
49.	8/26/2016	83.88	1	0	0	0

1 - 100 / 185

Fig. 5: Chart for Zero Memory Usage and Disk Throughput

The Fig. 5 Chart for Zero Memory Usage and Disk Throughput shows the chart for memory usage and disk throughput in which zero kb/sec data is used. Which can shows us we can delete these unused instances or we can stop these services on these dates so we can find out the unvisited instances dates efficiently.

IX. Conclusion and Future Research Directions

A. Summary of Key Findings

This research paper embarked on an analytical expedition to unravel efficiency and economies in Company X's cloud usage during the year 2016. Through rigorous data collection, exploratory data analysis, and data visualization using Google Looker Studio, we unearthed key insights into resource utilization patterns, temporal trends, and instances with zero activity.

The ascending order chart revealed resource-intensive instances that warrant optimized resource allocation. Conversely, instances with minimal resource consumption offer opportunities for cost savings through resource downsizing or termination. The heat map showcased temporal cloud usage patterns, enabling effective capacity planning for seasonal demands. Additionally, the chart highlighting instances with zero activity identified underutilized resources, indicating potential areas for optimization.

B. Implications for Cloud Usage Optimization

The findings of this research have substantial implications for optimizing Company X's cloud usage. By adopting data-driven decision-making, the organization can implement the following cost optimization strategies:

- *Optimal Resource Allocation:* Prioritize resource allocation to resource-intensive instances to maintain optimal performance.
- *Resource Downsizing:* Identify underutilized instances and reallocate resources to high-demand instances, reducing unnecessary costs.
- *Capacity Planning:* Develop proactive capacity planning strategies to align with seasonal resource demands, ensuring cost-efficient resource provisioning.
- *Instance Management:* Address instances with zero activity promptly, releasing unused resources and streamlining cloud services.
- *Continuous Monitoring:* Establish a robust monitoring system to detect changes in usage patterns, enabling timely adjustments to optimize costs.

C. Potential Future Research Directions

This research has opened avenues for potential future studies in cloud usage optimization and data analysis. The following research directions are suggested:

- *Longitudinal Analysis:* Conduct a longitudinal analysis over multiple years to examine how cloud usage trends evolve and identify long-term patterns.
- *Predictive Modeling:* Develop predictive models to forecast cloud usage trends, enabling proactive resource allocation and cost optimization.
- *Cloud Service Performance:* Explore the correlation between cloud service performance and resource utilization to optimize performance and cost-effectiveness.

- *Anomaly Detection:* Investigate anomaly detection techniques to identify abnormal resource usage and potential security threats in the cloud environment.
- *Cost-Benefit Analysis:* Perform a comprehensive cost-benefit analysis of implementing different cost optimization strategies to evaluate their effectiveness and impact on ROI.
- *Hybrid Cloud Optimization:* Extend the analysis to encompass hybrid cloud environments, addressing the unique challenges of optimizing resource allocation across multiple cloud platforms.

D. Conclusion

In conclusion, this research paper has delved into the depths of Company X's cloud usage data and illuminated essential insights through data analysis and visualization. By adopting Google Looker Studio's powerful capabilities, we have effectively communicated complex information, empowering Company X to make informed decisions for cloud usage optimization. The cost optimization strategies derived from this research offer a pathway to enhanced resource efficiency, improved performance, and substantial cost savings. By fostering a data-driven decision-making culture and embracing continuous improvement, Company X can navigate the ever-changing cloud landscape with adaptability and success. The discoveries and methodologies presented in this paper contribute valuable knowledge to the domain of cloud usage optimization, inspiring further research and advancements in cloud resource management. As the cloud computing landscape continues to evolve, the principles and insights derived from this research will guide organizations towards greater efficiency, competitiveness, and cost-effectiveness in cloud usage.

References

[1] V. Ramasamy, and S. M. Thalavai Pillai, "An effective HPSO-MGA optimization algorithm for dynamic resource allocation in cloud environment," *Cluster Comput*, vol. 23, no. 3, pp. 1711-1724, Sep. 2020, doi: 10.1007/s10586-020-03118-x.

[2] X. Wang, X. Wang, H. Che, K. Li, M. Huang, and C. Gao, "An intelligent economic approach for dynamic resource allocation in cloud services," *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 275-289, Jul. 2015, doi: 10.1109/TCC.2015.2415776.

[3] A. Bhardwaj, and C. Rama Krishna, "Impact of factors affecting pre-copy virtual machine migration technique for cloud computing," 2019. [Online]. Available: www.sciencedirect.comwww.materialstoday.com/proceedings

[4] I. Odun-Ayo, and F. Agono, "Cloud management optimization postgraduate view project cloud computing view project," 2018. [Online]. Available: https://www.researchgate.net/publication/332143932

[5] I. Iyoob, E. Zarifoglu, and A. B. Dieker, "Cloud computing operations research," *Service Science*, vol. 5, no. 2, pp. 88-101, Jun. 2013, doi: 10.1287/serv.1120.0038.

[6] E. Weintraub, and Y. Cohen, "Cost optimization of cloud computing services in a networked environment," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 4, 2015, doi: 10.14569/ijacsa.2015.060420.

[7] C. S. Pawar, and R. B. Wagh, "Priority based dynamic resource allocation in cloud computing," in *Proceedings - 2012 International Symposium on Cloud and Services Computing (ISCOS'12)*, IEEE Computer Society, 2012, pp. 1-6, doi: 10.1109/ISCOS.2012.14.

[8] U. K. Jena, P. K. Das, and M. R. Kabat, "Hybridization of meta-heuristic algorithm for load balancing in cloud computing environment," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2332-2342, Jun. 2022, doi: 10.1016/j.jksuci.2020.01.012.

[9] V. E. Jayanthi, R. Divya, and M. Jagannath, "A review on optimization approaches in cloud computing service," 2018. [Online]. Available: www.sciencepubco.com/index.php/IJET

[10] N. S. Kumari, K. Divya, C. Yadav, D. Asha, and M. Deepika, "A light weight secure data

sharing scheme for mobile cloud computing," *J Algebr Stat*, vol. 13, no. 3, pp. 3952-3956, 2022. [Online]. Available: https://publishoa.com

[11] M. Kumar, and S. C. Sharma, "Dynamic load balancing algorithm for balancing the workload among virtual machine in cloud computing," in *Procedia Computer Science*, vol. 115, pp. 322-329, 2017, doi: 10.1016/j.procs.2017.09.141.

[12] S. Afzal, and G. Kavitha, "Load balancing in cloud computing – A hierarchical taxonomical classification," *Journal of Cloud Computing*, vol. 8, no. 1, Dec. 01, 2019, doi: 10.1186/s13677-019-0146-7.

[13] Muhd. Junaid, A. Sohail, R. N. Bin Rais, A. Ahmed, I. Ali Khan, ... and N. Ejaz, "Modeling an optimized approach for load balancing in cloud," *IEEE Access*, vol. 8, pp. 173208-173226, 2020, doi: 10.1109/ACCESS.2020.3024113.

[14] P. Osypanka, and P. Nawrocki, "Resource usage cost optimization in cloud computing using machine learning," *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 2079-2089, 2022, doi: 10.1109/TCC.2020.3015769.

[15] Muhd. Jawad, Muhd. B. Qureshi, Muhd. U. S. Khan, S. M. Ali, A. Mehmood, ... and S. U. Khan, "A robust optimization technique for energy cost minimization of cloud data centers," *IEEE Transactions on Cloud Computing*, vol. 9, no. 2, pp. 447-460, Apr. 2021, doi: 10.1109/TCC.2018.2879948.