

Using Cloud AI for Email management through NLP at Accenture Ltd.

Tanvee Proag (2422814)

University of Mauritius

1 Background Study

Accenture Services Ltd. is a business consultancy firm with two distinct divisions - technology and operations. With 'Change' as its emblem, Accenture has always focused on how important it is to innovate and adapt in today's rapidly evolving business landscape. Therefore, it can be said that it is aligned with Accenture's values to implement NLP techniques through Cloud AI models which represent some of the most advanced artificial intelligence technologies currently available.

2 Problem Statement

At Accenture, the current email system used by employees introduces many challenges during the management of communications with clients. The cluttered email inboxes means that it is hard to navigate through the overwhelming volume of messages and as a consequence, employees spend a lot of time sorting through individual emails manually. This is inefficient and detracts from their productivity. Accenture currently uses Microsoft Outlook for Business for email management. It provides all the essential email management features and it integrates well with other Microsoft tools but due its lack of customization options and its heavy reliance on standard features, it is not as effective as it could be.

3 Proposed Solution

To improve efficiency and communication within the company, a customized and sophisticated email categorization system would improve efficiency and communication within Accenture as it would allow employees to classify emails by urgency and relevance. This would ensure that important tasks are prioritized effectively. For that, a cloud AI model for email management can be used. The NLP classification should perform text preprocessing, text embedding and pass them through a series of classification layers. The goal of the AI model is to reduce response time, improve customer satisfaction and operational efficiency.

4 Cloud Platform

For the cloud platform, both Amazon Web Services (AWS) and Google Cloud Platform (GCP) would be a good choice. AWS offers Amazon Comprehend for NLP tasks that would be needed such as entity recognition and sentiment analysis, SageMaker to use the emails for model building and training and deployment and AWS Lambda for serverless deployment. GCP provides Natural Language AI for NLP, AutoML and Vertex AI for custom machine learning models and Cloud Functions for serverless deployment. Natural Language AI could be used for sentiment analysis of the emails and classification into categories. Both AWS and GCP offer pricing that is flexible (the pay-as-you-go method) which means that the costs depend on the usage which is an advantage as since it does not need to be used all the time, Accenture gets to pay only when needed. Both AWS and GCP support autoscaling of models, that is, resources are only allocated according to demand. AWS has integrated features like EC2 and Lambda while GCP supports autoscaling through its Kubernetes Engine and Vertex AI. Amazon Elastic Compute Cloud (Amazon EC2) offers the broadest and deepest compute platform having more than 750 instances and choice of the latest processor, storage, networking, operating system and purchase model. This helps in matching the load of the NLP task of the whole organization. In terms of community support, both have developer communities and documentation, even though AWS has more third-party support resources. GCP focuses a lot on artificial intelligence and machine learning which could be advantageous for NLP-based tasks. Both cloud platforms are useful, powerful and well-equipped for the NLP tasks needed but out of the two, it might be better to use AWS as Accenture is already an AWS partner - the Accenture AWS Business Group (AABG) accelerates end-to-end adoption of AWS, while expanding enterprise-wide transformation and innovation safely, quickly and at scale.

5 Service Model

Each service model (Software as a Service - SaaS, Platform as a Service - PaaS, Infrastructure as a Service - IaaS) offers varying levels of control and infrastructure management. SaaS provides a fully managed solution where the provider handles everything, from infrastructure to application software. Users only need to interact with the application interface and have minimal control over the backend. It is easy to setup and scales well but offers limited customization as the consumer cannot modify the underlying infrastructure. PaaS offers a platform to build, train, and deploy AI models without managing the underlying hardware and operating system. The user can focus on model development while the provider manages the infrastructure. However, it might require some technical expertise. IaaS offers complete control over the infrastructure, but Accenture would be left with the responsibility of managing and configuring the AI models and software environment. It is significantly costlier and requires expertise. The most suitable service model would be PaaS which offers a platform to build,

train and deploy AI models without managing the underlying hardware and operating system. Then, Accenture can shift its focus to model development only as the provider already manages the infrastructure. PaaS provides a balanced solution between managing infrastructure and allowing customization which is particularly important for the flexibility to build and deploy the customized model of this case study. PaaS models offered by AWS (SageMaker) and GCP (Vertex AI) would allow more control over model building and its integration and would relieve infrastructure management issues. The advantage is that there will be no need to manage servers directly which proves PaaS to be cost-effective. Also, it is scalable as it allows expanding and reducing resources as is needed. However, it might require some technical expertise to build and deploy custom solutions. It normally costs higher than SaaS because of the customization and scaling but the custom NLP model for specialized email categorization might go beyond the scope of SaaS which makes PaaS a better choice of platform.

6 Deployment Model

The deployment model determines how the AI solution is hosted. In a public cloud deployment, the infrastructure and services are owned and operated by a third-party cloud provider (e.g., AWS, GCP, Azure) and shared across multiple organizations. They offer virtually limitless scalability and allow easy scaling up or down based on demand. However, it would be too risky for Accenture to allow the storing of sensitive data on shared infrastructure. Private cloud infrastructure is dedicated to a single organization, either hosted on-premises or by a third-party provider, and offers enhanced security and control over data and allows more customization than public cloud infrastructure. However, managing and maintaining a private cloud is more expensive than using a public cloud and it may require more expertise and dedicated IT resources. Hybrid models combine public and private clouds, allowing data and applications to be shared between them.

The preferred deployment model would be a hybrid cloud model as it allows sensitive data from the emails (e.g., client data, intellectual property, personally identified information, financial data etc.) to remain on a private cloud for security and public cloud resources can handle less-sensitive data processing. This helps with data governance, compliance (e.g., GDPR) and security concerns. For data privacy concerns, Accenture's global Client Data Protection (CDP) backed by security processes, policies and governance provides end-to-end security risk management and data security.

7 Integration

The problems that could crop up from integration should be avoided or a way to resolve them must be found. It must be ensured that the data formats align between the internal systems and the cloud AI platform. To do so, the emails

can be standardized and pre-processed and then fed to the NLP model. To protect sensitive data contained in the emails, the data must be encrypted and there should be a defined access control that is based on employees' roles. AWS and GCP already have strong enough security protocols but their compatibility with Accenture's encryption keys, access policies and compliance measures should be reviewed and implemented. Cloud provider libraries (e.g., AWS Boto3, GCP Client Libraries) help in integration as they allow efficient model training, deployment and API interactions with cloud services.

8 Deployment Strategy

Next, models are to be deployed as containerized applications using Docker or Kubernetes. This facilitates portability across environments. They can be easily scaled and managed, which works well with hybrid models. However, this would require expertise in the management of containerized applications. For a serverless architecture, AWS Lambda or GCP Cloud Functions can be deployed. It is a cost-effective choice and scales based on events which is great for intermittent usage. However, there is limited control over the runtime environment and performance may vary under heavy load. Models can be deployed on dedicated virtual machines in the cloud. This allows for full control over the environment and configurations which is useful for this resource-intensive application. However, it is less scalable and may require manual scaling which is more costly than serverless or containerized options.

9 Scalability and Resource Management

Considering auto-scaling, both AWS and GCP support features that allow instances to scale based on load. AWS Lambda scales functions based on the number of requests received. GCP offers managed instance groups that scale based on server load. Using load balancing services (e.g., AWS Elastic Load Balancing, GCP Load Balancer) evenly distributes traffic across instances, thus preventing any one instance from becoming a bottleneck. Autoscaling and serverless models reduce costs by scaling resources only when needed, which is beneficial as the workload is not constant but fluctuates in amount.

References

1. Amazon Web Services, Inc. (2019). Accenture and AWS. [online] Available at: <https://aws.amazon.com/accentureaws/> [Accessed 6 Nov. 2024].
2. Amazon Web Services, Inc. (n.d.). Amazon Comprehend – Features. [online] Available at: <https://aws.amazon.com/comprehend/features/> [Accessed 6 Nov. 2024].
3. AWS (2015). What Is Elastic Load Balancing? - Elastic Load Balancing. [online] Amazon.com. Available at: <https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/what-is-load-balancing.html> [Accessed 6 Nov. 2024].
4. AWS (2024). Amazon EC2. [online] Amazon Web Services, Inc. Available at: <https://aws.amazon.com/ec2/> [Accessed 6 Nov. 2024].
5. boto3.amazonaws.com. (n.d.). Boto3 documentation — Boto3 Docs 1.16.56 documentation. [online] Available at: <https://boto3.amazonaws.com/v1/documentation/api/latest/index.html> [Accessed 6 Nov. 2024].
6. cloud.google.com. (n.d.). Cloud Load Balancing — Google Cloud. [online] Available at: <https://cloud.google.com/load-balancing?hl=en> [Accessed 6 Nov. 2024].
7. Google Cloud. (n.d.). Google Kubernetes Engine (GKE). [online] Available at: <https://cloud.google.com/kubernetes-engine/?hl=en> [Accessed 6 Nov. 2024].
8. Google Cloud. (2018). Natural Language API Basics — Cloud Natural Language API — Google Cloud. [online] Available at: <https://cloud.google.com/natural-language/docs/basics> [Accessed 6 Nov. 2024].
9. Google Cloud. (2021). Cloud Client Libraries — Cloud APIs — Google Cloud. [online] Available at: <https://cloud.google.com/apis/docs/cloud-client-libraries> [Accessed 6 Nov. 2024].
10. Information Security at Accenture. (n.d.). Available at: <https://www.accenture.com/content/dam/accenture/final/capabilities/technology/technology-innovation/document/Accenture-Information-Security-at-Accenture.pdf> [Accessed 6 Nov. 2024].
11. www.accenture.com. (n.d.). Information Security — Accenture. [online] Available at: <https://www.accenture.com/us-en/services/technology/information-security> [Accessed 6 Nov. 2024].