Generative Adversarial Network (GAN) As A HPE GreenLake Service

Yogesh Choubey*, Sriram Ravishankar**, Soumya RR*, R, Harish*, Archana GS*, Anup Sahu Kumar*
NonStop Solutions* & CTO HPC/AI**
{yogesh.choubey, sriram.ravishankar, soumya.r-r, harish.r, archana.g-s, anup-kumar.sahu}@hpe.com

Abstract

Modern, forward-thinking organizations have become accustomed to agility and resilience to remain competitive in the information age. Virtualization, AI/ML, Cloud, and other transformative technologies are allowing enterprises to realize new operational efficiencies and business opportunities. The challenge now becomes on how to remain at the cutting edge while staying adaptable and secure, yet at the same time catering to existing needs and traditional modalities. GreenLake has capitalized on the above business opportunity by offering a self-service platform that marries the simplicity and agility of the public cloud with the security, governance compliance and performance benefits of an on-premises IT infrastructure, accessible through a seamless portal and complimented by a consumption-based pricing.

Although GreenLake offers a very robust solution, to keep up with the industry requirements and stay ahead of competitors in this space, multiple services could be offered through our platform which solves specific real-world problems. As seen recently with the emergence of GPT-4, data centric approaches to training large multimodal models are of utmost value, resulting in significant system performance improvements [1]. There is an acute shortage of quality real-world data to understand and train models that help predict results in certain fields like medicine, astronomy, genetics, and nuclear science. Generative Adversarial Networks (GANs) have shown their potential to generate realistic images, videos, and audio with the help of deep learning techniques [2]. Providing GAN as a service can be a practical solution for businesses that want to utilize GANs for various purposes without investing in expensive hardware and hiring specialized personnel. In this paper we would like to propose an approach for offering GAN architectures as a service on GreenLake.

Problem statement

Data has proven to be one of the most important components of R&D and innovation over the years. Since the advent of big data and analytics, data has been utilized to forecast outcomes and make key business decisions. Large amount of quality data is key for these accurate forecasts, resulting in high-quality process change especially in machine learning problems. Obtaining the right data is a time-consuming and often difficult endeavor due to unavailability, ethical, and legislative restrictions. For example, during the COVID-19 pandemic, data was very limited due to privacy and regulatory issues. Data augmentation solves the data limitation problem by synthesizing data from a given input and GAN has proven to be one of the most successful data augmentation techniques based on neural networks [3].

‘GAN As A Service’ (GANaaS) has the potential to benefit many industries, including healthcare, entertainment, e-commerce, and gaming. For instance, GANs can be used to generate synthetic medical images for training machine learning models for medical diagnosis. In the entertainment industry, GANs can be used to generate realistic images and animations for movies. They can be used to generate product images, which can enhance the customer experience in e-commerce. Lastly, in gaming, they are employed to generate environment textures and characters models as well.

Although GANs have shown impressive results, they require significant computational power such as high-performance CPUs and GPUs for training. With traditional hardware, the training process could take days or even weeks to compute, depending on the size of the model and the complexity of the data. Furthermore, training GANs requires expertise in deep learning, making it difficult for non-experts to use GANs for their applications.
Based on recent market study, the push for GAN services on cloud platforms in the industry has become evident. Figure 1 illustrates how these services are being accelerated to enhance efficiency and customer experience.

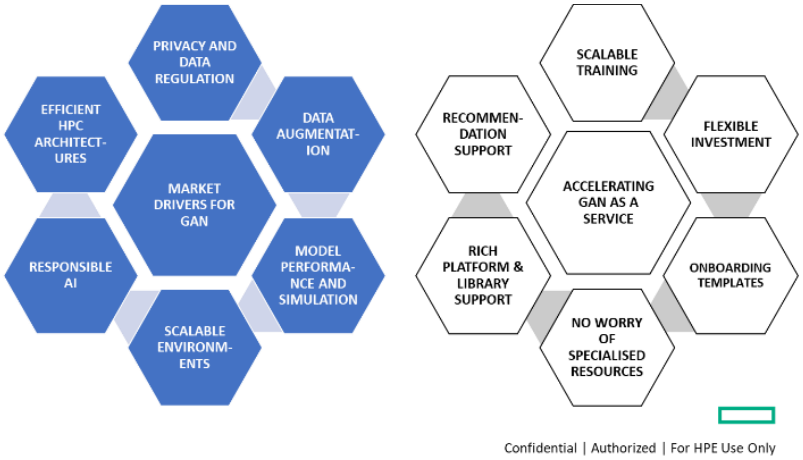
PRIVACY AND DATA REGULATION
EFFICIENT HPC ARCHITECT-URES
DATA AUGMENTAT-ION
MARKET DRIVERS FOR GAN
RESPONSIBLE AI
MODEL PERFORMA-NCE AND SIMULATION
SCALABLE ENVIRONM-ENTS
SCALABLE TRAINING
RECOMMEN-DATION SUPPORT
FLEXIBLE INVESTMENT
ACCELERATING GAN AS A SERVICE
RICH PLATFORM & LIBRARY SUPPORT
ONBOARDING TEMPLATES
NO WORRY OF SPECIALISED RESOURCES
Confidential | Authorized | For HPE Use Only

Fig 1. Market drivers for GAN and accelerating GANaaS

Our solution

To overcome the afore mentioned challenges and make GANs more accessible to users, we propose offering GAN as a service on our GreenLake platform. In this offering, users would be able to access tunable GAN frameworks (Bare GAN) and pre-trained GAN models via templates that are customized for specific user scenarios such as Image Resolution, data augmentation etc. Bare GAN can be customized in terms of their layers, hyperparameters, epochs based on the user's requirements whereas pre-trained models can directly be used from the available pool. The customized models can then be trained to generate the required data samples.

GANaaS approach would provide several benefits to users. First, it would eliminate the need for users to have deep learning expertise or access to powerful compute resources. Users would be able to generate high-quality, realistic samples using these customized models using HPEs underlying infrastructure. Second, offering GAN as a service on GreenLake would reduce the time and resources required to generate high-quality samples. Training a GAN can be a time-consuming and resource-intensive process on traditional hardware and this platform would help them accelerate this process. Third, the GANaaS model would allow users to generate samples for a wide range of applications, including image, audio, video synthesis, text generation, and other tasks. This would make GAN architectures more accessible to a wider range of users and enterprises, thus helping us to broaden our scope and avenues for business opportunity. The diagram stack shown in figure 2 visualizes our conceptualization of how GAN services would be offered with GreenLake.

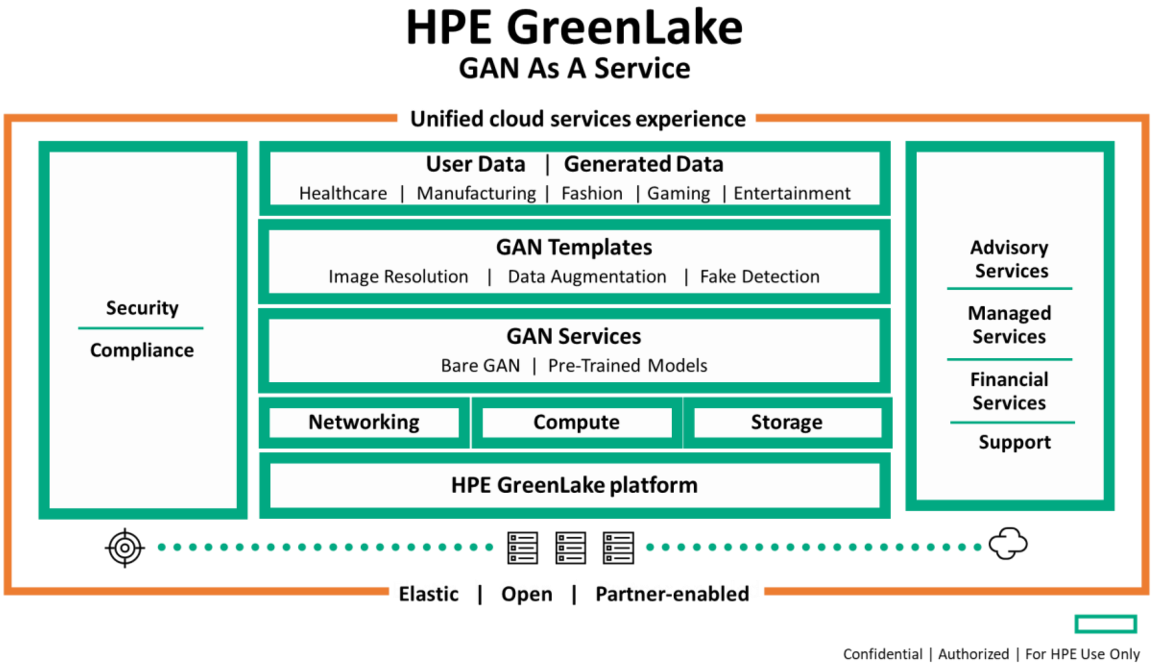
HPE GreenLake
GAN As A Service
Unified cloud services experience
Security
Compliance
User Data | Generated Data
Healthcare | Manufacturing | Fashion | Gaming | Entertainment
GAN Templates
Image Resolution | Data Augmentation | Fake Detection
GAN Services
Bare GAN | Pre-Trained Models
Networking
Compute
Storage
HPE GreenLake platform
Advisory Services
Managed Services
Financial Services
Support
Elastic | Open | Partner-enabled
Confidential | Authorized | For HPE Use Only

Figure 2. GAN as a service model

Evidence the solution works

Several examples and case studies suggest that GANaaS is a practical solution. Researchers from Carnegie Mellon University used GANs to generate synthetic training data for a machine learning model that could detect objects in aerial images. Researchers found that the GAN-generated data was as effective as real-world training data in improving the accuracy of the object detection model [4]. In another recent paper, researchers from Microsoft and the University of California, Berkeley used GANs to generate synthetic data for training a deep learning network that could recognize handwritten characters. Here also they found that the GAN-generated data was very close to real-world data and helped them in better prediction of the handwritten text [5]. These examples have served as inspiration and helped us understand how this service could be used effectively for a business use case or research. Porting such models to our Greenlake platform will demand a workflow that takes input data, customizes the GAN models for customer applications, train these models, track their progress, and generate the required data efficiently. The entire service could be offered as a metered product with the underlying MLOps frameworks complementing our GAN service and utilizing the necessary infrastructure. Figure 3 summarizes this process cycle.

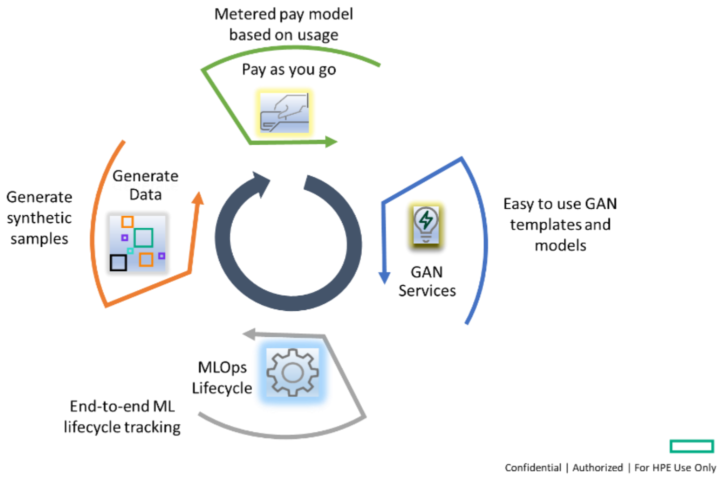

Figure 3. GAN As A Service Lifecycle

Competitive approaches

As GANs have started gaining popularity, several companies and organizations have developed competitive approaches to offering GAN as a Service. Microsoft has integrated AI-powered image-generation technology into its Bing search engine, Edge browser, and the Microsoft Designer, which has enhanced product efficiency. AWS's SageMaker, a cloud-based platform for building, training, and deploying machine learning models, includes built-in support for GANs, making it easy for users to train and deploy GAN models without the need for significant deep learning expertise. NVIDIA offers the NVIDIA DGX system, a powerful computing platform designed for deep learning and artificial intelligence workloads, includes built-in support and plugins for users to train and deploy GAN models with high performance and scalability. Hugging Face offers a cloud-based platform for natural language processing (NLP) tasks, including text generation using GANs. The success story of "HPE's Natural Language Processing platform for Question and Answer" [6] and the adoption of its enhanced version across a variety of business applications has served as a benchmark for our approach.

Current status

Initial experiments have been performed using the MNIST dataset on several public cloud platforms like AWS and GCP to validate their performance. We are currently exploring suitable ways to add these GANs to our platform and enabling the service to use the underlying MLOps frameworks that HPE's Ezmeral has already integrated with GreenLake. We are working on integrating certain open-source models to leverage AI concepts and help users also contribute to evolve the platform continuously. Information gathering to understand current user requirements and usage of these GANs to base a pricing model that would suit a wide range of customers is also underway.

Next steps

Offering pre-trained models on the platform for users to pick based on their use-case would be good to offer as a premium service so that users could access state-of-the-art models that generate samples rapidly without the need to train them. There are over 450 variants of GANs, each having a different architecture and end-use case. So, integrating such models with a recommendation system which helps users pick the most suited one for their application will be challenging.

Acknowledgements

Authors would like to acknowledge,

- “S, Madhusoodhana Chari” for guidance and inspiration.
- “Mopur, Satish Kumar” for mentorship on GANs.
- “Sudhakaran, Sonu” for enumerating the technical details of GreenLake platform.
- “C, Shiju” and “Jolly, Jinu Mariam” for their support in the surveying and exploring various GAN architectures.

References

[1] arXiv:2303.08774v3 [cs.CL] 27 Mar 2023.

[2] W. Li, "Image Synthesis and Editing with Generative Adversarial Networks (GANs): A Review," 2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4), London, United Kingdom, 2021, pp. 65-70, doi: 10.1109/WorldS451998.2021.9514052.

[3] Kiru, Muhammad & Belaton, Bahari & Chew, Xinying & Almotairi, Khaled & Hussein, Ahmad & Aminu, Maryam. (2022). Comparative analysis of some selected generative adversarial network models for image augmentation: a case study of COVID-19 x-ray and CT images. Journal of Intelligent & Fuzzy Systems. 43. 1-20. 10.3233/JIFS-220017.

[4] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli. Learning to Learn with Conditional Class Generative Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 9396-9405, June 2020.

[5] G. Zhang, X. Geng, and K. Keutzer. Generating High-Quality Synthetic Chinese Handwritten Characters via Generative Adversarial Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 2471-2478, February 2020.

[6] https://community.hpe.com/t5/advancing-life-work/hpe-s-natural-language-processing-platform-for-question-and/ba-p/7093877#.Y0kCtBxBzg8.

Keywords: GreenLake, GAN, As A Service, Cloud, Generative AI.