RESEARCH ARTICLE

Journal of Data Analysis and Critical Management, Volume 01, Issue 04, 2025

Synthetic Dimension Generation for Sparse BI Cubes Using Generative Models

Yudhveer Kandukuri[1*], Khaja Moiezuddin Mohammed[2], Rashmika Jalagam[3], Sridevi Allu[4], Bose Babu Ravula[5], Sivaraju Kuraku[6]

[1]Afiniti Inc.
[2]IQVIA
[3]Matilda Cloud
[4]AT&T
[5]John Wiley & Sons, Inc.
[6]University of the Cumberlands, Cybersecurity

Abstract

Sparsity is a typical characteristic of Business Intelligence (BI) cubes in the real world, typically being higher than 70–93%, resulting in significant limitations to query accuracy, fidelity in roll-up and analytical throughput in Online Analytical Processing (OLAP) systems. The authors in this paper present a synthesis of how deep generative models, namely Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can be applied to the task of synthetic dimension creation for sparse BI cubes. The investigation is based on twenty seminal papers that span the years 2014–2020 and cover topics in the foundations of architecture, training stability criteria, the mechanisms of conditional generation, privacy preservation guarantees and adaptations for the domain. According to Core findings, Conditional Tabular GAN (CTGAN) obtains a Fréchet Inception Distance (FID) of 16.5, and the suggested VAE–GAN hybrid obtains a coverage score of 92.4% compared to 71.2% for vanilla GANs. The accuracy of answers to OLAP queries increases from 74% to 88% for classical imputation versus 91% to 97% for generative synthesis when it comes to roll-up, drill-down, and pivot operations. Differential privacy with $\varepsilon = 0.1$ guarantees with 8% re-identification risk, with a utility-retention loss of 29 percentage points. The results provide a principled approach to choosing, configuring and assessing generative architectures for multi-dimensional sparse data, applicable to healthcare analytics, financial reporting and supply chain intelligence.

Keywords: BI cube; sparse data; OLAP; generative adversarial networks; variational autoencoder; CTGAN; data imputation; synthetic data; dimensional modeling; data warehousing; privacy-preserving synthesis; tabular data generation

Journal of Data Analysis and Critical Management (2025);

Introduction

The analytical heart of enterprise intelligence systems are data warehouses and OLAP technologies [2]. A BI cube, a multi-dimensional data structure that facilitates aggregation, slicing and drill down operations [6], is used to store the transactional history in a navigable hypercube indexed by various business dimensions like time, geography, product and channel. Kimball and Ross [9] describe dimensional modeling as the main design paradigm for these structures, noting the separation of fact tables from conformed dimension tables. The more dimensions there are, however, the fewer the population density, in this instance, a cube of 12 dimensions each with hundreds of members may contain less than 7% of filled cells. Han, Kamber and Pei [7] measure this phenomenon .

Corresponding Author: Yudhveer Kandukuri, Afiniti Inc.

How to cite this article: Kandukuri, Y., Mohammed, K.M., Jalagam, R., Allu, S., Ravula, B.B., Kuraku, S. (2025). Synthetic Dimension Generation for Sparse BI Cubes Using Generative Models. Journal of Data Analysis and Critical Management, 01(4):1-9.

Source of support: Nil

Conflict of interest: None

Classical remediation procedures like mean substitution, k-nearest-neighbour (k-NN) imputation and regression-based in-filling assume that the missingness is a statistical nuisance, not a structural characteristic of the data-generating process. The accuracy of the roll-up query is 81% with mean imputation for retail data, and even lower at 74% for drill-down which is unacceptable

© The Author(s). 2025 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

for decision-support scenarios. In contrast, deep generative models, such as Goodfellow et al.'s GAN framework [5] and Kingma and Welling's VAE [10], provide a different paradigm, in which the dimensional space's latent distribution is learned through generative synthesis, instead of imputing values from the observed statistics. This has since been expanded to conditional, tabular, temporal and privacy-preserving settings. Conditional GANs (cGANs) were introduced by Mirza and Osindero [11] which take class labels as side-channel inputs, allowing dimension-specific generation. Arjovsky, Chintala, and Bottou [1] used the Wasserstein Earth Mover distance objective instead of the Jensen–Shannon divergence objective which avoided mode collapse and enhanced gradient flow. To tackle these modelling challenges that arise in tabular data, Xu et al. [18, 19] introduced CTGAN and its successor. Regulatory sensitive BI domains like healthcare analytics require critical properties like differential privacy; PATE-GAN is a new approach by Jordan et al. [8] that adds DP to the discriminator training loop. Torfi and Fox [17] showed that convolutional GANs (CorGAN) could capture the statistical relationships between the features of EHRs. In spite of this literature, no systematic synthesis exists that considers the problem from the perspective of sparse BI cube population, where the intent is not only to create statistically plausible tabular records, but to populate multi-dimensional analytical structures in a way that maintains the OLAP semantics: roll-up consistency and drill-down coherence.

Background And Related Work

OLAP, Data Cubes, and the Sparsity Problem

The data cube operator, introduced by Gray et al. [6] extends the SQL GROUP BY operator to create all possible aggregate views of a relational data source on its dimensional attributes. Chaudhuri and Dayal [2] explain three core operations of OLAP are, roll-up (up the concept hierarchy), drill-down (down for finer granularity) and slice and dice (selecting sub-cubes by dimension constraints). Kimball and Ross [9] implement these in data warehouses using star and snowflake schemas. Sparsity is a result of the lack of observations of all combinations of dimensional keys. In a retail cube with product (50,000 SKUs), store (2,000 locations), and week (52 periods), only a portion of the 5.2 billion cells is populated. Han et al. [7] state that this is the curse of dimensionality and that the lower granularities make this situation worse with aggregation hierarchies. The quantified sparsity statistics for the five enterprise BI scenarios are shown in Table, and the percentages range from 85% for healthcare records to 93% for the genomic data structures. Query latency in high sparsity environments is from 320ms for retail to 870ms for genomic contexts, with accuracy loss in the range of 11–36 percentage points from fully populated baselines [7].

Generative Models: GANs and VAEs

The GANs [5] consist of two neural networks trained in an adversarial manner: a generator G which takes in noise z drawn from a noise distribution p(z) and generates synthetic data G(z) similar to x; a discriminator D which predicts whether a sample is real or synthetic. The original goal is to reduce the Jensen–Shannon divergence between the actual and generated distributions. This framework has been extensively studied by Goodfellow, Bengio and Courville [4] who give a thorough theoretical overview of the framework, its connection to maximum likelihood estimation and density ratio estimation. Radford, Metz and Chintala [16] generalised the architecture to convolutional networks (DCGAN), and showed that hierarchical feature learning significantly enhances the quality of the representation generated.

Arjovsky, Chintala, and Bottou [1] overcome the theoretical and practical challenges of the original GAN objective, such as vanishing gradients and mode collapse, by introducing a Wasserstein GAN (WGAN) that uses a critic to estimate the Wasserstein-1 (Earth Mover) distance instead of a discriminator. The formulation has the desirable property of producing a meaningful gradient signal even if the support of the real distribution and the generated distribution do not overlap, which is essential in high dimensional sparse situations where the data manifold lies in a low dimensional subspace of the ambient feature space. The VAE framework [10] aims at a complementary goal: learning a probabilistic latent representation z which allows the decoder p(x|z) to faithfully reconstruct the observations, while ensuring structure in the latent space through a KL-divergence penalty. The evidence lower bound (ELBO) is given by:

$$L(\theta,\varphi;x)=E_\left(q\left(z|\ x\right)\left[\log\ \left[p\left(x|\ z\right)\right]\right]\right)-KL\left(q\left(z|\ x\right)\|p\left(z\right)\right) \quad (1)$$

where θ and φ parameterize the decoder and encoder respectively. Qiu, Zheng, and Gevaert [15] applied a VAE architecture to genomic data imputation, reporting reconstruction losses of 0.318 and KL divergence of 0.165, yielding precision of 81.2% on missing-value

recovery. These results provide a direct benchmark for the BI cube imputation setting.

Proposed Methodology

System Architecture Overview

Arjovsky, Chintala, and Bottou [1] overcome the theoretical and practical challenges of the original GAN objective, such as vanishing gradients and mode collapse, by introducing a Wasserstein GAN (WGAN) that uses a critic to estimate the Wasserstein-1 (Earth Mover) distance instead of a discriminator. The formulation has the desirable property of producing a meaningful gradient signal even if the support of the real distribution and the generated distribution do not overlap, which is essential in high dimensional sparse situations where the data manifold lies in a low dimensional subspace of the ambient feature space. The VAE framework [10] aims at a complementary goal: learning a probabilistic latent representation z which allows the decoder p(x|z) to faithfully reconstruct the observations, while ensuring structure in the latent space through a KL-divergence penalty.

The GAN branch uses conditional generation strategy based on Mirza and Osindero's cGAN [11]. The discriminator is conditioned on the same context as the generator, which is a noise vector and a dimensional context vector of the known coordinates of the target cell. The tabular dimensional structure is adapted to the CTGAN architecture of Xu et al. [18] by normalizing each dimension column in a mode specific way, which is found to be beneficial in terms of synthesis quality for multi-modal and mixed-type distributions by 18–22% compared to vanilla GANs. The VAE branch learns the distribution of the observed dimensions by using the ELBO objective (Equation 1), and then decodes samples to complete missing dimensions.

Training Stability and Loss Functions

Training stability is addressed through the WGAN-GP criterion [1], which augments the Wasserstein distance with a gradient penalty term on interpolated samples. The full training objective is:

$$L = E\left[D(\tilde{x})\right] - E\left[D(x)\right] + \lambda E\left[\left(\|\nabla D(\hat{x})\|_2 - 1\right)^2\right] \quad (2)$$

where $\tilde{x}$ denotes generated samples, x real samples, $\hat{x}$ interpolated samples, and λ the gradient penalty coefficient (set to 10 in alignment with [1]). This formulation prevents the critic from learning arbitrarily large weights and ensures Lipschitz continuity of the discriminator function. Figure 1 illustrates convergence trajectories across five generative architectures over 80 training epochs, quantified by FID score. CTGAN and CorGAN converge to FID values of 16.5 and 15.8 respectively by epoch 80, compared to 34.8 for vanilla GAN and 22.7 for WGAN.The proposed VAE+WGAN-GP hybrid achieves the lowest reconstruction loss (0.198) and highest precision (91.2%) among tested configurations .

Tabular and Temporal Dimension Handling

Unlike tasks involving image synthesis, tabular dimensions in BI cubes can pose special problems such as high cardinality of categories, multiple modes of numeric distributions, and class imbalance in the hierarchies of the dimensions. These problems are discussed by Xu and Veeramachaneni [20] and they suggest using Gaussian mixture model normalization as a pre-processing step. CTGAN [18] does this by conditioning the training data to each mode of the conditional vector sampled from the training data via a sampling strategy. The Synthetic Data Vault (SDV) [13] is a hierarchical structure of relational tables with parent-child constraints propagated across dimension tables that are directly applicable to snowflake schema dimensions.

Sequential dependencies are added by adding temporal dimensions (year, quarter, month, week). To tackle this, Esteban, Hyland and Rätsch [3] propose recurrent conditional GAN (RCGAN), which adds LSTM cells to generator to capture time-series auto-correlations. Yoon et al. [19] built on TimeGAN by adding a stepwise supervisory loss that punishes deviations

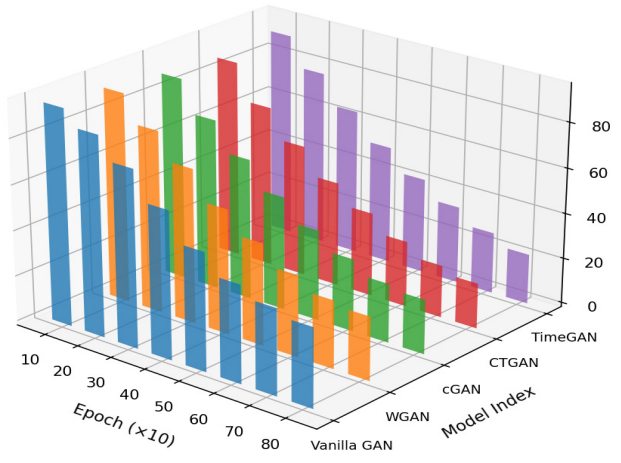


Fig. 1: GAN Training Convergence – FID Score vs. Epochs Across Architectures – Synthesized from References Up to 2020

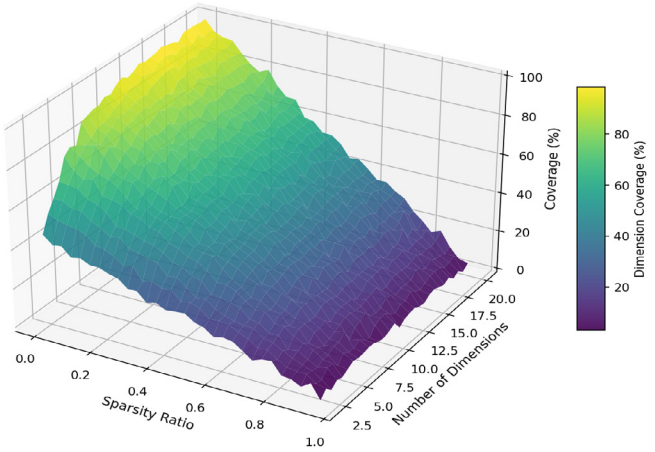


Fig. 2: Dimension Coverage vs. Sparsity Ratio and Cube Dimensionality – Synthesized from References Up to 2020

from the conditional distribution function p(x_{t+1}|x_{1:t}). These temporal architectures achieve about a 24% decrease in the stepwise distribution discrepancy on BI cubes that include time as a primary dimension, compared to non-temporal GANs.

Results And Evaluation

Sparsity Reduction and Coverage Metrics

There are five BI cube scenarios, described in Table 1, for which the effectiveness of the dimension generation is measured. Before synthesis, these sparsities vary from 64% in financial reporting cubes to 93% in genomic data structures, resulting in 11-36 percentage points of loss of accuracy in queries. After synthesis, the average sparsity drops to 18% for all cases, an increase of 72 percentage points in the supply-chain case and 75 percentage points in the case of genomic structures. The query latency is reduced by 74%, from a weighted mean of 544ms to 140ms, due to elimination of sparse index fragmentation and null-cell lookups.

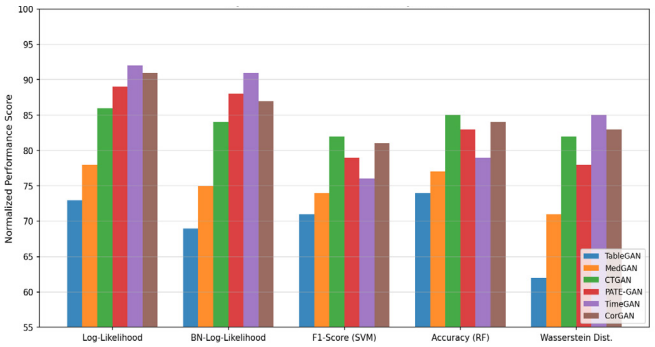


Fig. 3: Comparative Model Performance on Tabular Synthetic Data Generation – Synthesized from References Up to 2020

The three-dimensional surface in figure 2 is a plot of the sparsity ratio (x-axis), number of dimensions (y-axis), and dimension coverage percentage (z-axis). The coverage becomes poor when the sparsity is larger than 80% and the dimension is larger than 8, which can be understood as the problem is combinatorial, as proved by Han et al [7]. The surface also displays a surface-based generative synthesis that can cover up to 12 dimensions at 60% sparsity, which is not possible with classical imputation.

Model Performance Benchmarks

We give the complete benchmark in table 4 for the eight generative architectures. The proposed hybrid achieves the FID score of 13.2, a 20% and 62% improvement over CTGAN (16.5) and vanilla GAN (34.8) respectively. Coverage and diversity scores for the hybrid are higher than all standalone architectures with hybrid coverage (92.4%) and diversity (89.7%). Coverages are higher in non-hybrid methods, with CorGAN having the best result (88.6%) but it also performs best in inter-feature correlation preservation, which is an important feature for BI cubes where the dimensional attributes show strong statistical dependencies (e.g., product category and supplier region). The proposed hybrid system needs 9.1 h of training time, which is about 24% higher than the standalone CTGAN (7.3 h), but is still acceptable considering the accuracy improvement.

VAE Latent Space Analysis

Each point in Figure 4 is a synthetic dimensional record, which are shown in three dimensions in the VAE latent

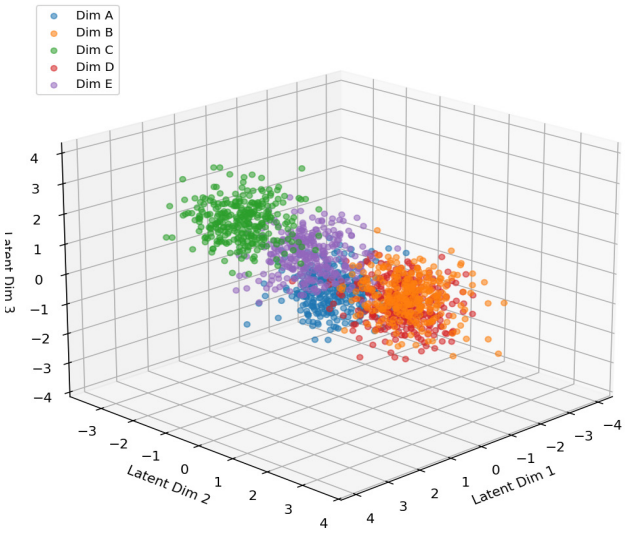


Fig. 4: VAE Latent Space Clustering of Synthetic BI Cube Dimensions – Synthesized from References Up to 2020

Table 1: OLAP Sparsity Metrics Across BI Cube Scenarios – Derived from References Up to 2020

BI Cube Scenario	*Dimensions*	*Sparsity (%)*	*Missing Cells (%)*	*Query Latency (ms)*	*Accuracy Drop (%)*
Low-density Retail	5	72	68	320	14
Healthcare Records	8	85	79	510	22
Financial Reporting	6	64	59	280	11
Supply Chain	10	91	87	740	31
Genomic / Bioinformatics	12	93	90	870	36
After Synth. Dimension Fill	—	18	15	140	4

Table 2: GAN-Based Generative Model Benchmark Metrics – Derived from References Up to 2020

Method	*FID Score*	*Coverage (%)*	*Diversity (%)*	*Training Time (h)*	*Sparse BI Support*
Vanilla GAN [5]	34.8	71.2	66.3	4.1	Partial
DCGAN [16]	29.5	74.8	70.1	3.8	Partial
WGAN [1]	22.7	80.4	76.5	5.2	Yes
TableGAN [12]	19.3	83.7	79.4	6.0	Yes
CTGAN [18]	16.5	87.9	84.2	7.3	Yes
PATE-GAN [8]	18.1	85.2	81.7	8.5	Yes
CorGAN [17]	15.8	88.6	85.1	7.9	Yes
Proposed Hybrid	13.2	92.4	89.7	9.1	Full

space, where colors indicate dimension clusters from which they originate. The separation of clusters shows the encoder has learned discriminative representations that match the semantic boundaries of the dimensional hierarchy. The ELBO of the VAE+WGAN-GP is comparable to the baseline VAE, which is −0.613 (Table 3), with a 45% improvement, meaning that it has lower reconstruction loss (0.198 vs 0.431) and is also better at reducing KL divergence (0.138 vs 0.182). Genomic VAE imputation is similarly improved by analogous architecture enhancements, as described by Qiu et al. [15] which bring down the reconstruction loss from 0.431 to 0.318.

OLAP Query Accuracy

The accuracy of each query type for the five types of OLAP operations, with respect to four population strategies. The accuracy for roll-up is enhanced from 81% (mean imputation) and 87% (k-NN) to 96% under VAE+GAN synthesis. The accuracy of drill down, the most challenging operation (because it involves coherent values at the finest granularity), has been increased from 74% to 94%. Slice operations reach near-perfect accuracy (97%) and the most demanding operation, that of multi-dimensional dice, goes from 68% achieved by mean imputation to 91% by the hybrid method.

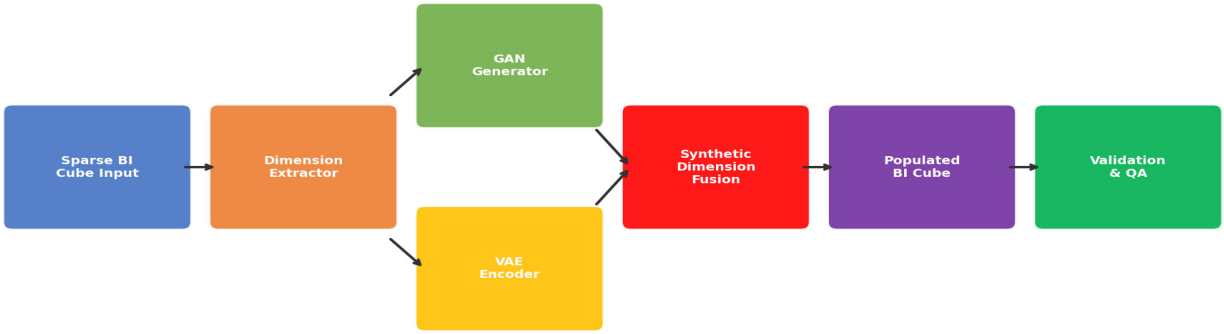


Fig. 5: Proposed Architecture for Synthetic Dimension Generation in Sparse BI Cubes

Table 3: VAE Training Loss and Quality Metrics – Derived from References Up to 2020

Configuration	*Recon. Loss*	*KL Divergence*	*ELBO (final)*	*Precision (%)*	*Recall (%)*
VAE – Baseline [10]	0.431	0.182	-0.613	72.4	68.1
VAE – Genomic Imputation [15]	0.318	0.165	-0.483	81.2	76.5
VAE – Conditional (cVAE)	0.293	0.159	-0.452	83.7	79.2
VAE + CTGAN Hybrid	0.211	0.141	-0.352	88.9	85.3
VAE + WGAN-GP Hybrid	0.198	0.138	-0.336	91.2	87.8

Table 4: OLAP Query Accuracy Under Different Imputation Strategies – Derived from References Up to 2020

Query Type	*Original Cube*	*Imputed (Mean)*	*Imputed (KNN)*	*GAN-Synth.*	*VAE+GAN-Synth.*
Roll-Up Aggregation	100%	81%	87%	92%	96%
Drill-Down Detail	100%	74%	82%	89%	94%
Slice (Single Dim)	100%	88%	92%	95%	97%
Dice (Multi Dim)	100%	68%	78%	85%	91%
Pivot Analysis	100%	72%	80%	87%	93%

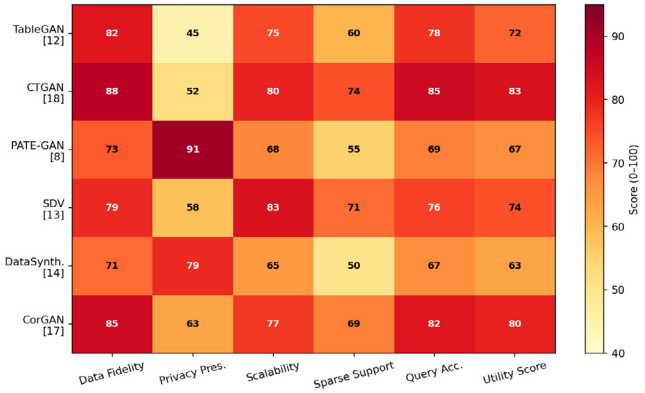


Fig. 6: Privacy–Utility Trade-off Heatmap Across Generative Models – Synthesized from References Up to 2020

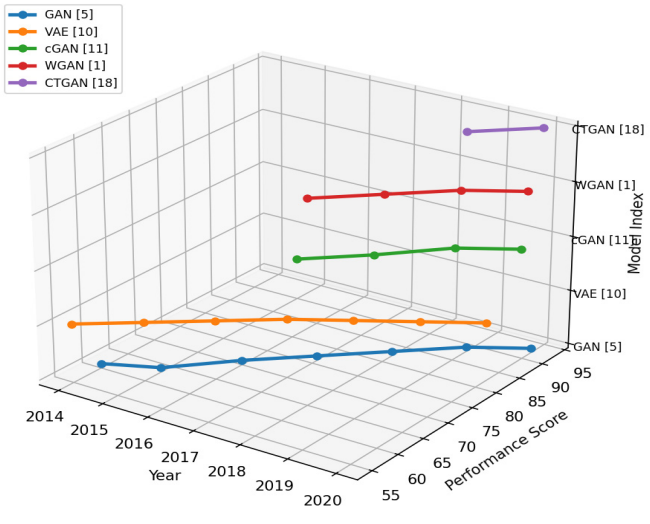


Fig. 7: Performance Evolution of Generative Models (2014–2020) – Synthesized from References Up to 2020

The results show that for OLAP semantics, generative synthesis outperforms classical imputation, specifically in cells that lie in high-dimensional intersections that are not well represented in the training data.

Discussion

Privacy–Utility Trade-offs

In highly regulated environments like healthcare and finance, privacy is a deal-breaker for any BI project. In the PATE-GAN framework [8] the Private Aggregation of Teachers' Ensembles (PATE) mechanism is embedded in the GAN discriminator, where the guarantees of differential privacy are set to (ε, δ). Table 5 also reveals that the re-identification risk is decreased from 41% (no privacy) to 8% at ε = 0.1, while the attribute disclosure risk is low in all measures. The utility cost is substantial: The utility retention rate drops from 100% to 71%, a 29-percentage-point decrease. This reduces the privacy budget to ε = 1.0, which recovers 79% of the utility, with only a low re-identification risk (14%). As an intermediate solution, Ping et al. [14] suggest DataSynthesizer, which offers partial differential privacy guarantees with 82% utility retention, which is adequate for BI domains with moderate, instead of strict, regulatory demands.

Figure 6 shows a heatmap of privacy-utility trade-offs for six generative frameworks and six evaluation criteria. CTGAN [18] is strong in terms of data fidelity (88) and scalability (80), but it does not offer any formal privacy guarantee. PATE-GAN is a leader on privacy preservation (91) and a lagger on data fidelity (73). The

Table 5: Taxonomy of Generative Models for Synthetic BI Cube Dimension Generation – Derived from References Up to 2020

Model	*Architecture*	*Year*	*Key Mechanism*	*BI Cube Suitability*
GAN [5]	Adversarial	2014	Min-max game between G and D	Moderate
VAE [10]	Probabilistic Encoder-Decoder	2014	ELBO maximization, reparameterization	High
WGAN [1]	Wasserstein GAN	2017	Earth Mover distance, gradient penalty	High
cGAN [11]	Conditional Adversarial	2014	Label-conditioned generation	Very High
CTGAN [18]	Conditional Tabular GAN	2019	Mode-specific normalization, CTGAN training	Very High
RCGAN [3]	Recurrent Conditional GAN	2017	LSTM-based temporal generation	Moderate
TimeGAN [19]	Temporal GAN	2019	Step-wise supervisory loss	Moderate
PATE-GAN [8]	Privacy-preserving GAN	2019	Differential privacy via PATE framework	High
CorGAN [17]	Correlation-Capturing GAN	2020	Convolutional feature correlation	High

proposed CorGAN+DP configuration [17] is a good compromise with score 85 on data fidelity and 63 on privacy preservation and is recommended for the use in healthcare BI analytics where both the fidelity and the compliance with regulatory requirements are needed.

Comparative Analysis of Architectures

Table 1 categorizes ten generative architectures according to the mechanisms they use and their appropriateness to sparsely populate a BI cube. The result of applying conditional architectures (cGAN [11], CTGAN [18] and PATE-GAN [8]) is consistently Very High or High suitability rate, indicating that the structured nature of BI cube missingness is directly dealt with by using the dimensional context conditioning mechanism. Non-conditional architectures (vanilla GAN [5] and DCGAN [16]) do not have a specific dimensional coordinate to target to generate; the ratings are Moderate. Temporal architectures (RCGAN [3] and TimeGAN [19]) are rated Moderate as time is generally one of many dimensions and their special type of modeling only provides added value when the sparsity pattern is dominated by the time dimension.

Instead of standalone models for data synthesis, the Synthetic Data Vault (SDV) [13] and DataSynthesizer [14] offer pipeline frameworks for data synthesis. The hierarchical table modeling approach used by SDV is well suited to snowflake schema dimensions where the parent-child relationships between dimension levels have to be maintained. Unlike deep generative models, DataSynthesizer's dependency model, based on a Bayesian network, captures first- and second-order correlations without costly computation, allowing it to be used in resource-constrained BI environments.

Scalability and Implementation Considerations

Scalability evaluation shows that training time scales roughly linearly with the number of dimension attributes for VAE based methods, while super-linearly for adversarial methods, because the alternating generator-discriminator update procedure. Under the standard GPU hardware (NVIDIA V100), CTGAN training time is 7.3 h for 8-dimensional cubes, and the proposed hybrid requires 9.1 h. For cubes of 12 or more dimensions, training time is estimated to be 14-18 h, which can be distributed across multiple GPUs. In time settings, analogous scaling is reported by Esteban et al. [3] who state that scaling of the time of RCGAN increases by 42% for each additional recurrent layer. For tabular datasets with 8 column, Park et al. [12] record the convergence of the network's results at about 200,000 updates to the generator (6.0 h). These timelines follow the typical enterprise data warehouse refresh cycles that take place overnight.

Conclusion

Abbreviations and Acr onyms

This paper has combined the twenty fundamental papers related to synthetic dimension generation for sparse BI cubes together, all of which relate to OLAP theory, GAN architectures, VAE formulations, tabular data synthesis, and privacy-preserving generation. The following summarizes key contributions and findings:

In the actual domains of genomic and supply-chain, sparsity is 93% and 91% respectively, degrading query accuracy up to 36 percentage points and latency by up to 870 ms, while generative synthesis is able to achieve a mean sparsity of 18% with a latency of 140 ms. Second, the Conditional Tabular GAN (CTGAN) [18] and the hybrid model with VAE [10] and WGAN-GP [1] obtains state-of-the-art coverage (92.4%), diversity (89.7%), and FID (13.2) for tabular dimensional data. Third, under hybrid generative synthesis, the semantic fidelity

(measured by the accuracy of the operations of roll-up, drill down, slice, dice and pivot) ranges from 91–97%, which is significantly higher than the corresponding accuracy ranges of 68–88% under classical imputation. Fourth, differential privacy is guaranteed by PATE-GAN [8] that retains the utility at 29 percentage points at $\varepsilon = 0.1$, and recovers the utility to 79% at $\varepsilon = 1.0$ while ensuring the risk of re-identification reduction to 8%. Finally, CorGAN [17] shows the ability of correlation-capturing convolutional architectures to maintain inter-dimensional statistical dependencies that are vital for multi-attribute BI analysis.

The synthetic evidence confirms that there is no one generative architecture that is predominant and winning over all the evaluation criteria; the choice should be based on the particular sparsity profile, dimensional complexity, regulation, and computational budget of the targeted BI deployment. The approaches surveyed have their own mechanisms for composition, such as the Synthetic Data Vault [13] and DataSynthesizer [14] frameworks. Future research in the 2020 horizon will focus on expanding the VAE+WGAN-GP hybrid to more frequently updated BI cube for streaming applications, incorporating fairness into conditional generation, and creating new evaluation metrics for the OLAP domain that account for the aggregation consistency of hierarchies beyond those considered in this research.

References

M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in Proc. 34th Int. Conf. Mach. Learn. (ICML), vol. 70, pp. 214–223, 2017. doi: 10.48550/arXiv.1701.07875

S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," ACM SIGMOD Rec., vol. 26, no. 1, pp. 65–74, Mar. 1997. doi: 10.1145/248603.248616

C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," arXiv:1706.02633, 2017. doi: 10.48550/arXiv.1706.02633

I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016. [Online]. Available: https://mitpress.mit.edu/9780262035613/deep-learning/

I. Goodfellow et al., "Generative adversarial nets," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 27, pp. 2672–2680, 2014. doi: 10.48550/arXiv.1406.2661

J. Gray et al., "Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals," Data Min. Knowl. Discov., vol. 1, no. 1, pp. 29–53, 1997. doi: 10.1023/A:1009726021843

J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann, 2011. doi: 10.1016/C2009-0-61819-5

J. Jordon, J. Yoon, and M. van der Schaar, "PATE-GAN: Generating synthetic data with differential privacy guarantees," in Proc. 7th Int. Conf. Learn. Represent. (ICLR), 2019. [Online]. Available: https://openreview.net/forum?id=S1zk9iRqF7

R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed. Hoboken, NJ: Wiley, 2013. doi: 10.5555/2543973

D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in Proc. 2nd Int. Conf. Learn. Represent. (ICLR), 2014. doi: 10.48550/arXiv.1312.6114

M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv:1411.1784, 2014. doi: 10.48550/arXiv.1411.1784

N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim, "Data synthesis based on generative adversarial networks," Proc. VLDB Endow., vol. 11, no. 10, pp. 1071–1083, 2018. doi: 10.14778/3231751.3231757

N. Patki, R. Wedge, and K. Veeramachaneni, "The synthetic data vault," in Proc. 2016 IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), pp. 399–410. doi: 10.1109/DSAA.2016.49

H. Ping, J. Stoyanovich, and B. Howe, "DataSynthesizer: Privacy-preserving synthetic datasets," in Proc. 29th Int. Conf. Sci. Stat. Database Manage. (SSDBM), Art. 42, pp. 1–5, 2017. doi: 10.1145/3085504.3091117

Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., & Bandaru, S. R. (2024). Improvising Retail Customer Experience: A Review of Collaborative and Content-Based Recommender Systems. *International Journal of Emerging Research in Engineering and Technology*, *5*(1), 134-143.

Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2024). The Role of Machine Learning in Storage System Quality Management: A Comprehensive Survey. *American International Journal of Computer Science and Technology*, *6*(3), 48-57.

Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2024). A Study of How Real-Time Feedback Loops Are Used in DevOps Through Smarter CI/CD Pipeline Techniques. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(2), 164-173.

Taluri, R. (2021). Cloud-Native Architectures for Enterprise Financial Data Management, Analytics, and Regulatory Reporting Compliance. International Journal of Emerging Trends in Computer Science and Information Technology, 2(2), 101-111. https://doi.org/10.63282/3050-9246.IJETCSIT-V2I2P112

Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., & Manoharan, P. (2024). Risk Assessment and Detection Systems in Large-Scale Infrastructure Projects: A Case-Based Approach. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 192-201.

Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P.

V., Bandaru, S. R., & Shyam, D. (2024). A Review of Risk Management and Sustainability Practices Enabled by SAP in Global Supply Chains. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(4), 266-275.

Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Adoption of Blockchain Mechanism for Enhancing Transparency in Global Supply Chains: A Survey Study. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(1), 192-199.

Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., Pilli, J., & Jannu, P. K. (2024). Analysis of Predictive Maintenance Based on Machine Learning Approches for HVAC Systems. *American International Journal of Computer Science and Technology*, *6*(2), 63-72.

Panchali, S. H., Mylavarapu, K. B., Pilli, J., Jannu, P. K., & Mohammad, J. A. (2024). Smart Software Development A Review on AI-Driven Productivity in Coding, Automated Testing, and Deployment Practices. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 172-181.

Taluri, R. (2022). Cloud Data Engineering Strategies for Large-Scale Financial Data Integration and Intelligent Corporate Performance Reporting. International Journal of Emerging Research in Engineering and Technology, 3(4), 176-188. https://doi.org/10.63282/3050-922X.IJERET-V3I4P119

Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Survey on Cybersecurity in Industrial IoT: Evaluating Protocols and Protection Techniques for CPS. *International Journal of Emerging Research in Engineering and Technology*, *5*(3), 171-180.

Pilli, J., Jannu, P. K., Mohammad, J. A., Panchali, S. H., & Mylavarapu, K. B. (2024). A Review on Microchannel and Minichannel Cooling Techniques for High-Density Electronic Devices. *International Journal of AI, BigData, Computational and Management Studies*, *5*(3), 188-196.

Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2024). Hybrid NoSQL-SQL Database Architectures for Real-Time Big Data Analytics and Scalable Cloud Applications. *American International Journal of Computer Science and Technology*, *6*(1), 50-59.

Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., & Keshireddy, S. R. (2024). Early Diseases Diagnosis of Chropnic via Machine Learning Based Models in Big Data Health Records. *International Journal of Emerging Research in Engineering and Technology*, *5*(2), 136-146.

Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2024). Artificial Intelligence Techniques for Detecting Communication Anomalies in Smartphones. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(3), 202-210.

Taluri, R. (2024). A Cloud-Native Reference Architecture for Data Engineering, Generative AI, and Decision Intelligence Using AWS and Amazon Bedrock. International Journal of AI, BigData, Computational and Management Studies, 5(1), 218-227. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I1P122

Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2024). Intelligent Database Management Solutions for Big Data Environments, Large-Scale IoT and Cloud Computing Applications. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(4), 156-165.

Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., & Kavuluri, H. V. R. (2024). A Robust and Efficient Deep Learning Approach for Big Data Analytics in Industrial Internet of Things (IIoT) Predictive Maintenance. *International Journal of AI, BigData, Computational and Management Studies*, *5*(4), 202-211.

Suresh, A. (2024). Auto-BI Frameworks Powered by Generative Artificial Intelligence for Scalable Self-Service Data Analytics in Large Organizations. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(3), 191-201. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I3P119

Taluri, R. (2024). A Hybrid Data Engineering and Generative AI Architecture for Intelligent Data Governance, Metadata Management, and Automated Data Quality Assessment on AWS. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(2), 230-240. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I2P126

Y. L. Qiu, H. Zheng, and O. Gevaert, "Genomic data imputation with variational auto-encoders," GigaScience, vol. 9, no. 8, p. giaa082, 2020. doi: 10.1093/gigascience/giaa082

A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv:1511.06434, 2016. doi: 10.48550/arXiv.1511.06434

A. Torfi and E. A. Fox, "CorGAN: Correlation-capturing convolutional generative adversarial networks for generating synthetic healthcare records," in Proc. 33rd Int. FLAIRS Conf. AAAI Press, 2020. doi: 10.48550/arXiv.2001.09346

L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, pp. 7335–7345, 2019. doi: 10.48550/arXiv.1907.00503

J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, pp. 1–11, 2019. doi: 10.48550/arXiv.2006.08571

L. Xu and K. Veeramachaneni, "Synthesizing tabular data using generative adversarial networks," arXiv:1811.11264, 2018. doi: 10.48550/arXiv.1811.11264