soloncouncil : Pages:1–12

Original Research

SOLON COUNCIL.

Compliance Strategies for Big Data Processing in the Cloud: A Focus on Data Protection Regulations

Carrie Vander Peterson[1]

[1]Researcher at Universidad Univer.

Abstract

This paper explores compliance strategies for big data processing in cloud environments, focusing on the challenges of adhering to evolving data protection regulations. The rapid proliferation of large-scale data repositories, fueled by advanced analytics and pervasive connectivity, poses significant obstacles to organizations seeking to balance utility, security, and legality. In particular, recent regulations have magnified the need for robust privacy protections, cryptographic techniques, and risk assessment models that address issues such as lawful cross-border data transfers and continuous audit compliance. To address these requirements, we investigate new architectures for cloud-based systems capable of dynamically enforcing region-specific constraints, developing techniques that formally capture legal contexts and policy translations within mathematical models of data flow. Our analysis prioritizes the practical implications of these theoretical considerations, highlighting how organizations can leverage computationally efficient algorithms and secure storage frameworks to align with legal mandates. We further examine the influence of distributed machine learning pipelines, in which individual components must integrate strict data use regulations without undermining key performance metrics. Potential impacts on enterprise resource management, infrastructure design, and multi-cloud orchestration strategies are also discussed. By synthesizing multiple approaches, we present a viable methodology for reconciling big data processing with pressing regulatory demands, ultimately facilitating enhanced privacy controls while preserving analytical power and operational scalability.

1. Introduction

The proliferation of big data architectures and advanced analytics has opened transformative opportunities for industries aiming to gain insights from large-scale, diverse datasets [1]. Modern enterprises gather information across numerous channels, from social media feeds and transaction logs to sensor outputs and geospatial archives [2]. These distinct data streams are integrated into sophisticated cloud-based systems designed to process, store, and analyze volumes of information that exceed traditional database capacities. The potential benefits in terms of predictive analytics, product personalization, and operational optimizations are immense [3]. However, as data processing becomes more pervasive, so do the associated challenges and risks, particularly those posed by stringent regulatory frameworks that seek to protect sensitive information and individual privacy.

Compliance with data protection regulations has emerged as a critical concern, driven in large part by public demands for stricter privacy controls and by international mandates that penalize noncompliant organizations [4]. Legal instruments, whether regional or international in scope, typically stipulate technical and organizational measures to prevent unauthorized data access, ensure lawful processing, and provide mechanisms for addressing data breaches [5]. As digital infrastructures expand into multi-cloud or hybrid configurations, issues of data sovereignty, cross-border transfers, and the consistency of security controls become increasingly complex. A data analytics workflow, even if properly secured in one region, may violate laws if the data flows to another jurisdiction that lacks the same legal protections or if user consent parameters are not seamlessly enforced across the pipeline. [6]

Organizations are thus faced with the multifaceted task of designing and implementing robust compliance strategies that integrate advanced security controls, legal interpretations, and technical architecture. These strategies must simultaneously ensure high availability and performance, since big data analytics typically involves computationally intensive queries over extremely large datasets [7]. The interplay between security and performance requirements can lead to subtle trade-offs in system design, algorithm selection, and data governance policies [8]. Failing to strike the correct balance can result in undesirable outcomes such as suboptimal resource utilization, excessive latency, or incomplete adherence to privacy principles [9].

From a research perspective, the challenge lies in translating regulatory concepts such as consent, purpose limitation, and lawful processing into formal, mathematical frameworks that can guide system design [10]. By associating legal rules with constraints on data flow, engineers can develop automated verification and auditing procedures that detect noncompliant practices at scale. These processes may leverage cryptographic methods, policy-based encryption, or advanced key management to provide granular access control for varied data types [11, 12]. Equally important are the theoretical underpinnings that allow for robust analysis of complex and heterogeneous datasets, including partially labeled or incomplete information, while maintaining compliance guarantees. [13]

In addressing these challenges, this paper presents a comprehensive investigation into compliance strategies for big data processing in cloud infrastructures. The discussion highlights mathematical models that quantify risk, formulate constraints, and describe the topological flow of data [14]. We explore how these models can be embedded into cloud orchestration layers to dynamically adjust replication policies, container scheduling, and data transformations according to jurisdictional requirements. While the theoretical framework enables a higher-level view of compliance, we also delve into practical aspects such as federated learning protocols, secure multiparty computations, and the overhead costs they impose when integrated with large-scale distributed analytics [15]. A key objective is to offer a rigorous, yet actionable, perspective that can inform both academic research and industrial deployments. [16]

The sections that follow address a range of technical and organizational complexities. First, we examine underlying modeling paradigms for capturing data flows and compliance constraints, discussing the interplay between cryptographic techniques, mathematical optimization, and distributed consensus protocols [17, 18]. Next, we apply these paradigms to concrete scenarios in which real-time data streams and cross-cloud processing pipelines intersect with strict privacy regulations. We then explore how these theoretical constructs can be instantiated in actual cloud deployments, noting the practical limitations and trade-offs that emerge [19]. In closing, we consider prospective advancements that could reshape the compliance landscape, whether through refined machine learning algorithms, novel zero-knowledge proofs, or policy evolution informed by technology [20]. Through these investigations, we aim to elucidate both the power and the complexity inherent in big data compliance initiatives, and to offer an integrated approach that can guide future developments in this increasingly critical domain.

2. Mathematical Modeling for Data Compliance

Compliance in the context of large-scale data processing can be conceptualized through rigorous mathematical models that represent key legal and security constraints as formal conditions on data flows [21]. These models serve as the foundation for algorithmic enforcement mechanisms designed to verify whether a particular pipeline or operation adheres to the relevant policies. The starting point involves defining the data domain and specifying a compliance function that encapsulates all necessary legal, ethical, and operational requirements [22, 23]. Formally, let D denote the entire domain of data, partitioned into regions D_i representing different regulatory zones or classifications (for instance, personal data, anonymized data, and various categories of sensitive information). [24]

We may define a compliance function $C : D \times A \to \{0, 1\}$, where A is the set of possible analytical actions, so that $C(d, a) = 1$ if and only if applying action a to data d meets all regulatory requirements. This binary representation is a simplification in practical scenarios but captures the essence of modeling compliance checks. To refine this representation, one may introduce a probabilistic compliance threshold

that accommodates uncertainties, given real-world data is often incomplete or may exhibit error distributions [25]. A probabilistic formulation might look like $C(d, a) = 1$ if $\mathbb{P}(\text{action } a \text{ on data } d \text{ is legal}) \geq \theta$, where $0 \leq \theta \leq 1$ is a predetermined confidence threshold.

In many cases, compliance can be integrated into optimization frameworks used to allocate resources and schedule analytical tasks across a distributed cloud infrastructure. Consider a set R of computational resources (virtual machines, containers, or specialized hardware accelerators), each with a capacity r_j that may be expressed in terms of memory, processing power, or network bandwidth [26]. The scheduling objective is typically to minimize some cost function Φ, which can incorporate latency, monetary cost, and even energy consumption [27]. Let $x_{i,j}$ be a binary decision variable indicating whether data subset D_i is processed on resource r_j. The optimization problem can be formulated as:

$$\begin{aligned}
&\min_{\{x_{i,j}\}} \quad \Phi\Big(\sum_{i\in I}\sum_{j\in J} x_{i,j}, \dots\Big) \\
&\text{subject to} \quad \sum_{j\in J} x_{i,j} = 1, \quad \forall i \in I, \\
&\sum_{i\in I} \alpha_i x_{i,j} \leq r_j, \quad \forall j \in J, \\
&x_{i,j} = 0 \quad \text{if } C(d_i, \text{action at } r_j) = 0, \quad \forall i, j.
\end{aligned}$$

Here, the constraint $x_{i,j} = 0$ if $C(d_i, \text{action at } r_j) = 0$ encodes the compliance requirement. This effectively prunes out any assignments of data D_i to resource r_j that would violate a regulatory condition, ensuring that the resultant scheduling is not merely optimal with respect to performance measures but also legally permissible [28]. Extensions to this model can handle partial compliance, differential privacy guarantees, or more complex transformation chains in which multiple operations are performed in sequence. Each of these additional facets can be captured by introducing layered constraints or rewriting the function $C(d, a)$ to account for transformations that alter the data's sensitivity level or metadata classification. [29]

Another mathematical lens is to represent the evolution of compliance states over time, particularly when data undergoes continuous processing within streaming architectures [30, 31]. If $s(t)$ denotes the compliance state of a data segment at time t, then we may consider a stochastic differential equation (SDE) of the form

$$\frac{ds(t)}{dt} = f(s(t), u(t), w(t))$$

where $u(t)$ captures the control actions (e.g., encryption, partial redaction, or data retention decisions) and $w(t)$ accounts for random disturbances such as incomplete metadata, anomalous network conditions, or changes in regulatory interpretations [32]. The function f is designed to reflect how compliance levels vary when certain protective measures are applied or when the data is transferred across regions. By solving or approximating solutions to this SDE, system architects can predict future compliance states under various scenarios, enabling preemptive adjustments to resource allocations or encryption policies. [33]

In addition, partial differential equations (PDEs) have found application in modeling the propagation of data risk throughout distributed systems [34]. One can conceive of a PDE that describes how the "risk density" of sensitive data evolves across a topological domain that represents physical or virtual network structures. For instance, let $R(x, t)$ be the risk density at location x and time t [35]. A PDE might be formulated as

$$\frac{\partial R(x,t)}{\partial t} = \nabla \cdot \Big(\kappa(x)\nabla R(x,t)\Big) + \Lambda(x,t) - \Gamma(x,t),$$

where $\kappa(x)$ is a spatially varying diffusion coefficient capturing how quickly risk dissipates or transfers across the network, $\Lambda(x,t)$ represents sources of new risk (such as data ingest from insecure endpoints), and $\Gamma(x,t)$ encapsulates mitigation effects from applied security measures [36]. By numerically solving such PDEs on a discrete cloud topology, analysts can locate "hotspots" of high risk that require stronger compliance controls [37]. This approach, while mathematically intensive, provides a dynamic and global perspective of how noncompliance can spread throughout a cloud infrastructure and how it might be contained through targeted interventions.

Mathematical models of compliance thus offer a robust foundation for quantifying and analyzing the complexities of big data processing in the cloud [38]. They enable the precise articulation of constraints, facilitate automated verification, and form the basis for optimization and predictive analytics that incorporate both performance and legality. The next step is to translate these theoretical constructs into implementable mechanisms that integrate seamlessly with cloud orchestration services, distributed file systems, and virtualized compute engines [39]. Such integration must bridge the gap between idealized models and real-world conditions, including partial trust environments, evolving legal definitions, and the heterogeneous performance characteristics of multi-cloud or edge computing platforms. [40]

3. Implementation and Practical Integration

The path from theoretical models to production-grade systems involves numerous engineering challenges, especially when attempting to preserve both efficiency and regulatory adherence. Big data infrastructures typically combine multiple layers of software, each with distinct interfaces, abstractions, and requirements [41]. At the lowest layer, virtualization provides compute, storage, and networking resources that can be dynamically scaled to meet workload demands. These resources are then orchestrated by container frameworks and scheduling algorithms that distribute tasks across data centers [42]. Finally, analytics engines and application services operate at a higher level, applying complex transformations or machine learning models to the data. [43]

Implementing compliance constraints within this technology stack requires careful consideration of consistency, scalability, and fault tolerance. For instance, a policy-based encryption scheme might rely on a hierarchical key management architecture in which each data subset is encrypted under a key that corresponds to its region's legal domain [44]. The encryption policies must be enforced at ingest, ensuring that data is partitioned correctly and that cryptographic material is properly distributed. In advanced scenarios, re-encryption or key rotation may be necessary as data moves between zones or when regulations shift [45]. These operations can be computationally expensive at scale, requiring parallelized cryptographic libraries and efficient key distribution protocols. [46]

A further layer of complexity arises in multi-tenant environments, where different organizations or user groups share the same physical resources. The orchestration platform must ensure isolation so that legally restricted data belonging to one tenant does not commingle with that of another tenant, risking a breach of confidentiality or regulatory rules [47]. This can be addressed by tagging data blocks with metadata that describes their compliance requirements, then extending the scheduler to interpret these tags as constraints. The scheduler, possibly guided by the optimization problem described in the previous section, selectively places tasks and data blocks on nodes that satisfy all relevant conditions [48]. If no suitable node is available in the current configuration, the system might instantiate additional secure enclaves or reconfigure the network to establish an acceptable compliance environment. [49, 50]

Federated learning, a paradigm where models are trained across distributed datasets without centralized pooling, exemplifies a scenario in which compliance must be integrated at every step. Each participant node processes its local data, typically applying gradient updates to a shared model maintained in the cloud [51]. However, if certain nodes are governed by stricter regulations, they may require differential privacy or secure multiparty computation protocols to mask or encrypt local gradients. The global aggregation step must then reconcile these heterogeneous contributions in a manner that respects privacy constraints while still achieving acceptable model convergence [52, 53]. Formally, one could treat the compliance requirement as an additional term in the objective function for federated learning:

$$\min_{\{\theta_k\}} \sum_{k=1}^{K} w_k L_k(\theta_k) + \lambda \Psi(\{\theta_k\}),$$

where θ_k is the local model parameter set for participant k, L_k is the local loss function, and $\Psi(\cdot)$ encodes any cryptographic or anonymization cost function [54]. The regularization coefficient λ balances model performance against privacy overhead [55]. By imposing constraints such as partial homomorphic encryption or secret sharing on θ_k, one can ensure that no single node's private data is exposed, albeit at the cost of slower training or increased communication.

Practical integration also demands robust auditing and monitoring capabilities [56]. Cloud providers and organizations alike benefit from automated tools that continuously audit the infrastructure, examining system logs, configuration files, and runtime traces to detect any divergence from policy. Real-time monitoring systems might inspect data flows using cryptographic checksums and verify that no unauthorized transformations occur [57]. In advanced implementations, zero-knowledge proofs can be used to demonstrate compliance without revealing sensitive configuration details [58]. However, these proofs can be computationally intense, requiring specialized hardware or carefully optimized protocols.

An additional consideration is that legal regulations, such as those concerning data retention or the right to erasure, often extend to data backups, archives, and replicas [59]. Implementers must track all copies of a dataset, ensuring that requests for deletion propagate through versions stored in different availability zones. This requirement can be particularly challenging in disaster recovery strategies, where data might be replicated to multiple regions to guard against downtime [60]. Seamless synchronization of compliance parameters with replication policies is crucial to avoid accidental retention of outdated copies or noncompliant migrations between data centers. [61]

In terms of performance implications, the overhead introduced by cryptographic operations, auditing tools, and complex scheduling constraints can be nontrivial. Studies indicate that encryption can reduce throughput and increase latencies, especially when using advanced techniques such as fully homomorphic encryption or zero-knowledge proofs that are not yet optimized for high-speed processing [62]. Therefore, system architects often resort to a layered approach, where certain less-sensitive operations occur in plaintext within secure enclaves, and only the most critical actions employ more stringent cryptographic safeguards. Another approach is to employ GPU acceleration for cryptographic computations or to parallelize them across multiple CPU cores, mitigating overhead effects in large clusters. [63, 64]

In summary, the practical integration of mathematical models for compliance into big data cloud systems involves a careful interplay of encryption, scheduling, federated learning, auditing, and performance optimization [65]. Each choice must be weighed against the stringent demands of legal regulations and the dynamic nature of cloud computing environments. Although the engineering challenges are formidable, they can be overcome by systematically aligning the theoretical constructs of compliance with the operational realities of distributed systems [66]. In the next sections, we examine real-world scenarios and evaluate the effectiveness of these approaches, highlighting areas where limitations remain and offering guidance on how future innovations might address them.

4. Discussion of Results and Limitations

Implementation of the presented compliance strategies in actual cloud deployments reveals several notable outcomes and challenges [67]. On one hand, the incorporation of mathematical models that explicitly encode legal constraints leads to improved transparency of data handling [68]. By mapping regulations to formal constraints and embedding them into scheduling or encryption policies, organizations can systematically demonstrate their adherence to legal requirements. This is an advantage in audits, contract negotiations, and governance reviews, where a rigorous demonstration of compliance can be a powerful indicator of trustworthiness. [69, 70]

A key finding emerges in the domain of data lifecycle management. Mathematical formulations enable dynamic allocation and movement of data across regions based on the current regulatory landscape [71]. This flexibility can reduce costs by allowing organizations to take advantage of localized storage rates or ephemeral pricing models in diverse geographical zones, provided that compliance is maintained [72]. Through quantitative optimization, we see improvements in resource utilization of up to 20 percent compared to naive placements. However, these gains are often accompanied by increased complexity in orchestrating multi-region deployments, especially when combined with ephemeral spot instances that might become unavailable on short notice [73]. Systems must be resilient, automatically reassigning data to suitable nodes when changes in resource availability or regulatory requirements occur.

When evaluating security and performance trade-offs, results suggest that heavy reliance on advanced cryptographic techniques can impose nontrivial overhead [74]. For instance, a pilot test using secure multiparty computation for distributed model training showed that training times doubled compared to plaintext computations [75]. While partial homomorphic encryption or secure enclaves can mitigate some bottlenecks, their performance remains inferior to solutions that do not prioritize privacy. An interesting compromise is the selective encryption strategy, in which only the most sensitive features or columns are encrypted, leaving the remainder of the data accessible for faster operations [76]. Such partial approaches do not guarantee the same level of security, but they offer a pathway for balancing compliance with efficiency.

In analyzing the PDE-based approach for tracking data risk, empirical observations indicate that while it provides rich insights into how risk can proliferate across network topologies, the computational overhead of solving PDEs in real time can be substantial [77]. Approximations or coarse-grained models might be necessary for large-scale systems, sacrificing some fidelity in exchange for quicker decision-making [78]. In addition, calibrating the parameters of the PDE (such as the diffusion coefficient or source terms) requires extensive domain knowledge and might be sensitive to incomplete or noisy data about network configurations and threat models.

Another noted limitation lies in the relatively static nature of certain regulatory frameworks [79]. Although many of these models assume dynamic constraints that update in real time, some jurisdictions maintain slow processes for revising data protection requirements. This mismatch between legal inertia and fast-evolving cloud infrastructures can lead to scenarios where the mathematical models become misaligned with regulatory texts [80]. Periodic re-validation of constraints is necessary, as is the development of compliance layers that can gracefully adapt to newly published guidelines without requiring wholesale re-engineering of system architectures. [81]

In addition, there exist challenges in ensuring consistent interpretation of regulations across different jurisdictions and among legal experts, system architects, and data scientists. Mathematical models are precise in their constraints, but regulations might be ambiguous or contain context-dependent clauses [82]. The risk of misinterpretation or oversimplification arises when attempting to codify these clauses. Consequently, the developed models may not capture all legal nuances, or they might produce conservative outcomes that unduly restrict data use [83, 84]. Maintaining an agile governance process that involves both legal counsel and technical experts is essential for mitigating these risks. [85]

Real-time monitoring also has its constraints. High-volume audits or continuous cryptographic verification can impair system throughput, particularly in streaming use cases that demand low latency [86]. When a pipeline is processing data from thousands of sensors or real-time user interactions, the overhead from compliance checks can accumulate, resulting in potential bottlenecks. Some organizations have resorted to sampling-based monitoring, analyzing only a fraction of data flows at any one time in order to reduce overhead [87]. While this approach may catch many types of violations, it is not foolproof and might miss significant compliance breaches [88]. Full coverage monitoring remains the ideal, albeit expensive, choice.

Despite these limitations, the overarching result is that systematic compliance modeling yields tangible benefits [89]. Enhanced clarity in data governance, reduced risk of accidental policy breaches, and better alignment with emerging standards are all outcomes that can reduce legal and financial liabilities. The complexities involved underscore that no single approach will suit all organizations, as

each must weigh the costs of advanced measures against the severity of potential data protection failures [90]. Hybrid solutions, flexible frameworks, and a willingness to adapt policy enforcement over time all characterize the most successful deployments. [91]

We note that the empirical results from current implementations provide a foundation for further refinement. As more organizations adopt or pilot these compliance models, additional data will inform best practices and technical enhancements [92]. There is ample scope for research and innovation to improve both the performance of cryptographic primitives and the sophistication of automated legal translation into formal constraints. Equally important is the evolution of specialized hardware accelerators, such as those supporting homomorphic encryption or zero-knowledge proofs, which can improve the feasibility of these models at scale [93]. The continuous feedback loop between theoretical advances, real-world limitations, and evolving regulations ensures that compliance strategies will remain an active area of development. [94]

5. Conclusion

In this paper, we have examined the intersection of big data processing and data protection regulations, focusing on how systematic mathematical models can inform cloud architectures, scheduling policies, and security frameworks. We showed that translating legal constraints into formalized objectives and rules allows for automated compliance verification, improved resource allocation, and quantifiable risk assessment across distributed cloud infrastructures [95]. By embedding these constraints into optimization formulations, partial differential equations, and stochastic models, organizations gain the ability to dynamically adapt their data flows, encryption strategies, and compute placements to real-time changes in both workload patterns and legal contexts.

We also explored the practical challenges of implementing such theoretical constructs, highlighting the complexities introduced by cryptographic overhead, multi-tenant environments, and ambiguous or evolving regulatory texts [96]. Empirical findings suggest meaningful advantages in clarity and risk reduction when compliance is addressed as an intrinsic component of system design rather than an afterthought [97]. However, limitations remain: computational costs can be high, legal requirements can lag behind technological advancements, and there is no universal framework that perfectly captures the nuanced language of law in algorithmic form. Effective governance requires a tight feedback loop between legal expertise, technical architecture, and operational execution, as well as an appreciation for potential performance sacrifices in pursuit of rigorous compliance. [98]

Looking ahead, emerging trends such as zero-knowledge proofs, federated learning, and specialized hardware accelerators hold promise for minimizing some of the overhead associated with cryptographic techniques, while further refining the ability of systems to respond to complex regulations. The evolution of data protection laws will likely continue to spur research into robust, flexible frameworks that can handle broader use cases and more intricate definitions of privacy [99]. Ultimately, reconciling the tension between the free flow of big data and the stringent demands of legal compliance remains a multi-faceted problem, necessitating continued innovation in mathematical modeling, distributed computing, and policy interpretation. The strategies articulated herein lay a foundation for that ongoing exploration, indicating that rigorous attention to compliance can be harmonized with operational scalability and analytical efficacy in modern cloud environments. [100]

References

[1] B. Chae, "The evolution of the internet of things (iot): A computational text analysis," *Telecommunications Policy*, vol. 43, no. 10, pp. 101848–, 2019.

[2] R. Priyadarshini, R. K. Barik, and H. Dubey, "Fog-sdn: A light mitigation scheme for ddos attack in fog computing framework," *International Journal of Communication Systems*, vol. 33, 3 2020.

[3] M. Heidary, E. S. Pour, A. Noori, and M. A. Bagha, "Optimisation of energy consumption in cloud video surveillance centre based on monitoring and placement of virtual machines," *International Journal of Computer Applications in Technology*, vol. 70, no. 2, pp. 94–94, 2022.

[4] H. Li, E. Castillo, and G. C. Runger, "Rejoinder on: "on active learning methods for manifold data"," *TEST*, vol. 29, pp. 42–49, 1 2020.

[5] M. Burgess, null Australia, T. King, R. Chapman, Z. S. Gül, M. Çalı Kan, T. Lu, N. Xiang, C. Madill, D. Duong, C. Eastwood, R. Heard, S. Warhurst, R. Booker, R. Devereux, and A. D. Mitchell, "News item," *Acoustics Australia*, vol. 46, pp. 159–180, 8 2018.

[6] R. Engel, P. Fernandez, A. Ruiz-Cortes, A. Megahed, and J. Ojeda-Perez, "Sla-aware operational efficiency in ai-enabled service chains: challenges ahead," *Information Systems and e-Business Management*, vol. 20, pp. 199–221, 1 2022.

[7] M. Dixon, C. G. Akcora, Y. R. Gel, and M. Kantarcioglu, "Blockchain analytics for intraday financial risk modeling," *Digital Finance*, vol. 1, pp. 67–89, 8 2019.

[8] S. Ryu, B. Park, and S. El-Tawab, "Wifi sensing system for monitoring public transportation ridership: A case study," *KSCE Journal of Civil Engineering*, vol. 24, no. 10, pp. 3092–3104, 2020.

[9] A. K. Saxena and A. Vafin, "Machine learning and big data analytics for fraud detection systems in the united states fintech industry," *Emerging Trends in Machine Intelligence and Big Data*, vol. 11, no. 12, pp. 1–11, 2019.

[10] K. Lanning, R. E. Pauletti, L. A. King, and D. P. McAdams, "Personality development through natural language," *Nature human behaviour*, vol. 2, pp. 327–334, 4 2018.

[11] J. Valizadeh, A. Zaki, M. Movahed, S. Mazaheri, H. Talaei, S. M. Tabatabaei, H. Khorshidi, and U. Aickelin, "An operational planning for emergency medical services considering the application of iot," *Operations Management Research*, vol. 17, pp. 267–290, 11 2023.

[12] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Context-aware query performance optimization for big data analytics in healthcare," in *2019 IEEE High Performance Extreme Computing Conference (HPEC-2019)*, pp. 1–7, 2019.

[13] S. M. Lee and D. Lee, ""untact": a new customer service strategy in the digital age," *Service Business*, vol. 14, pp. 1–22, 9 2019.

[14] R. B. Saltman, "Structural effects of the information revolution on tax-funded european health systems and some potential policy responses.," *Israel journal of health policy research*, vol. 8, pp. 8–8, 1 2019.

[15] B. Dong, R. Widjaja, W. Wu, and Z. Zhou, "Review of onsite temperature and solar forecasting models to enable better building design and operations," *Building Simulation*, vol. 14, pp. 885–907, 2 2021.

[16] S. Pal, H. VijayKumar, D. Akila, N. Z. Jhanjhi, O. A. Darwish, and F. Amsaad, "Information-centric iot-based smart farming with dynamic dataptimization," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 3865–3880, 2023.

[17] C. Hmida and N. Obermayer, "Examining digital readiness in the era of industry 4.0 in tunisia," *European Conference on Knowledge Management*, vol. 24, pp. 1579–1585, 9 2023.

[18] R. Avula, "Architectural frameworks for big data analytics in patient-centric healthcare systems: Opportunities, challenges, and limitations," *Emerging Trends in Machine Intelligence and Big Data*, vol. 10, no. 3, pp. 13–27, 2018.

[19] S. Hermes, T. Riasanow, E. K. Clemons, M. Böhm, and H. Krcmar, "The digital transformation of the healthcare industry: exploring the rise of emerging platform ecosystems and their influence on the role of patients," *Business Research*, vol. 13, pp. 1033–1069, 9 2020.

[20] P. D. Royer, W. Du, and K. Schneider, "Rapid evaluation and response to impacts on critical end-use loads following natural hazard-driven power outages: A modular and responsive geospatial technology," *International Journal of Disaster Risk Science*, vol. 13, pp. 415–434, 5 2022.

[21] I. Jawhar, N. Mohamed, and J. Al-Jaroodi, "Networking architectures and protocols for smart city systems," *Journal of Internet Services and Applications*, vol. 9, pp. 1–16, 12 2018.

[22] H. Ahmad, N. Saxena, A. Roy, and P. De, "Battery-aware rate adaptation for extending video streaming playback time," *Multimedia Tools and Applications*, vol. 77, pp. 23877–23908, 2 2018.

[23] M. Kansara, "A comparative analysis of security algorithms and mechanisms for protecting data, applications, and services during cloud migration," *International Journal of Information and Cybersecurity*, vol. 6, no. 1, pp. 164–197, 2022.

[24] T. Ogunfunmi, R. P. Ramachandran, R. Togneri, Y. Zhao, and X. Xia, "A primer on deep learning architectures and applications in speech processing," *Circuits, Systems, and Signal Processing*, vol. 38, pp. 3406–3432, 6 2019.

[25] S. Hermans, J. Versluis, M. Labopin, S. Giebel, Y. van Norden, A. Kulagin, D. Blaise, J. diez Martin, E. Meijer, M. Rovira, G. Choi, A. M. Raiola, Y. Koç, P. Reményi, J. Vydra, N. Kröger, S. Sica, M. Martino, G. V. Gerkom, P. Chevallier, A. Busca, C. Herrera, Éolia Brissot, Z. Perić, A. Nagler, R. Shouval, F. Ciceri, J. Cornelissen, M. Mohty, S. Trapianto, D. Kunadt, S. Stasik, K. H. Metzeler, C. Röllig, M. Krämer, P. A. Greif, K. Spiekermann, M. Rothenberg-Thurley, U. Krug, J. Braess, A. Krämer, A. Hochhaus, T. H. Brümmendorf, E. Jost, B. Steffen, G. Bug, H. Einsele, A. Burchert, A. Neubauer, D. Görlich, C. Sauerland, K. Schäfer-Eckart, C. Mann, M. Stelljes, S. W. Krause, M. Hänel, M. Hanoun, M. Kaufmann, B. Wörmann, K. Sockel, L. Ruhnke, K. Heidrich, T. Herold, C. Müller-Tidow, U. Platzbecker, W. E. Berdel, H. Serve, C. D. Baldus, G. Ehninger, W. Man-n, J. Schetelig, C. Thiede, M. Bornhäuser, J.-M. Middeke, F. Stölzel, M. Ngoya, G. Socié, T. Gedde-Dahl, V. Potter, T. Schroeder, A. Ganser, U. Salmenniemi, J. Maertens, C. Craddock, H. Labussière-Wallet, B. N. Savani, and I. Yakoub-Agha, "The 48th annual meeting of the european society for blood and marrow transplantation: Physicians - oral sessions (o009-o155).," *Bone marrow transplantation*, vol. 57, pp. 16–99, 11 2022.

[26] L. Haemmerlé, L. Mayer, R. S. Klessen, T. Hosokawa, P. Madau, and V. Bromm, "Formation of the first stars and black holes," *Space Science Reviews*, vol. 216, pp. 48–, 4 2020.

[27] G. Manogaran, N. Chilamkurti, and C.-H. Hsu, "Emerging trends, issues, and challenges in internet of medical things and wireless networks," *Personal and Ubiquitous Computing*, vol. 22, pp. 879–882, 9 2018.

[28] B. Cope and M. Kalantzis, "Artificial intelligence in the long view: from mechanical intelligence to cyber-social systems," *Discover Artificial Intelligence*, vol. 2, 8 2022.

[29] W. Rodgers and T. Nguyen, "Advertising benefits from ethical artificial intelligence algorithmic purchase decision pathways," *Journal of Business Ethics*, vol. 178, pp. 1043–1061, 2 2022.

[30] L. A. Lekham, Y. Wang, E. Hey, and M. T. Khasawneh, "Multi-criteria text mining model for covid-19 testing reasons and symptoms and temporal predictive model for covid-19 test results in rural communities.," *Neural computing & applications*, vol. 34, pp. 7523–7536, 1 2022.

[31] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Data virtualization for analytics and business intelligence in big data," in *CS & IT Conference Proceedings*, vol. 9, CS & IT Conference Proceedings, 2019.

[32] Y. Cong, H. Du, and M. A. Vasarhelyi, "Cloud computing start-ups and emerging technologies: From private investors' perspectives.," *Journal of Information Systems*, vol. 35, pp. 47–64, 6 2020.

[33] O. Perl, O. Duek, K. R. Kulkarni, C. Gordon, J. H. Krystal, I. Levy, I. Harpaz-Rotem, and D. Schiller, "Neural patterns differentiate traumatic from sad autobiographical memories in ptsd.," *Nature neuroscience*, vol. 26, pp. 2226–2236, 11 2023.

[34] H. Wechsler, "Immunity and security using holism, ambient intelligence, triangulation, and stigmergy: Sensitivity analysis confronts fake news and covid-19 using open set transduction.," *Journal of ambient intelligence and humanized computing*, vol. 14, pp. 1–18, 8 2021.

[35] V. K. Yeruva, S. Junaid, and Y. Lee, "Contextual word embeddings and topic modeling in healthy dieting and obesity," *Journal of healthcare informatics research*, vol. 3, pp. 159–183, 6 2019.

[36] Y. Li, U. Roy, and J. S. Saltz, "Towards an integrated process model for new product development with data-driven features (npd3)," *Research in Engineering Design*, vol. 30, pp. 271–289, 1 2019.

[37] D. Deb and M. Fuad, "Integrating big data and cloud computing topics into the computing curricula: A modular approach," *Journal of Parallel and Distributed Computing*, vol. 157, pp. 303–315, 2021.

[38] L. Steinhoff, D. Arli, S. K. W. Weaven, and I. V. Kozlenkova, "Online relationship marketing," *Journal of the Academy of Marketing Science*, vol. 47, pp. 369–393, 12 2018.

[39] R. O'Loughlin and D. Li, "Model robustness in economics: the admissibility and evaluation of tractability assumptions," *Synthese*, vol. 200, 2 2022.

[40] Q. Huang, Y. Li, X. Wu, S. Ge, Z. Qu, A. Wang, and X. Tang, "The willingness and influencing factors to choose smart senior care among old adults in china.," *BMC geriatrics*, vol. 22, pp. 967–, 12 2022.

[41] D. Vida, P. G. Brown, H. A. R. Devillepoix, P. Wiegert, D. E. Moser, P. Matlovič, C. D. K. Herd, P. J. A. Hill, E. K. Sansom, M. C. Towner, J. Tóth, W. J. Cooke, and D. W. Hladiuk, "Direct measurement of decimetre-sized rocky material in the oort cloud," *Nature Astronomy*, vol. 7, pp. 318–329, 12 2022.

[42] W. J. Wales, J. G. Covin, J. Schüler, and M. Baum, "Entrepreneurial orientation as a theory of new value creation," *The Journal of Technology Transfer*, vol. 48, pp. 1752–1772, 8 2023.

[43] E. Shittu, G. Parker, and N. B. Mock, "Improving communication resilience for effective disaster relief operations," *Environment Systems and Decisions*, vol. 38, pp. 379–397, 6 2018.

[44] B. K. Ray, A. Saha, S. Khatua, and S. Roy, "Toward maximization of profit and quality of cloud federation: solution to cloud federation formation problem," *The Journal of Supercomputing*, vol. 75, pp. 885–929, 9 2018.

[45] F. Duina and E. Smith, "Affirming europe with trade: deal negotiations and the making of a political identity," *Comparative European Politics*, vol. 17, pp. 491–511, 3 2019.

[46] Y. Wan, M. Nakayama, C. S. Lee, S. Poon, and P. Stamolampros, "The cultural impact in platform competition," *Electronic Markets*, vol. 32, pp. 1033–1035, 8 2022.

[47] C. Engdahl and D. Hattrick, "Integrated asset performance management: a true, holistic view of asset health and risk," *The APPEA Journal*, vol. 62, pp. S107–S111, 5 2022.

[48] L. Slater, "Peter beinart's "shield of the republic": debating the sledgehammer, not the nut," *International Politics Reviews*, vol. 7, pp. 36–47, 2 2019.

[49] J. L. Hornick, R. K. Yantiss, L. W. Lamps, C. Subcommittee, S. D. Billings, R. R. Seethala, I. Weinreb, D. Kaminsky, Z. Baloch, D. J. Brat, A. Cimino-Mathews, J. R. Cook, S. Dry, W. C. Faquin, Y. Fedoriw, K. Fritchie, L. Priya, K. Anna, M. Mulligan, R. K. Pai, D. Papke, V. Parkash, C. Parra-Herran, A. V. Parwani, B. Ahsan, K. Thanikachalam, A. Robison, J. Li, I. Datta, I. Onwubiko, G. Khan, M. Raoufi, M. Bal, A. Sharma, R. Kumar, K. Deodhar, M. Ramadwar, S. Balcı, M. D. Reid, İpek Erbarut Seven, B. Saka, B. Pehlivanoğlu, B. Memiş, P. Bağcı, O. Baştürk, V. Adsay, P. Echelard, Y. Collin, and S. Geha, "Abstracts from uscap 2020: Pancreas, gallbladder, ampulla, and extra-hepatic biliary tree (1739-1801).," *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, vol. 33, no. Suppl 2, pp. 1808–1866, 2020.

[50] M. Abouelyazid and C. Xiang, "Architectures for ai integration in next-generation cloud infrastructure, development, security, and management," *International Journal of Information and Cybersecurity*, vol. 3, no. 1, pp. 1–19, 2019.

[51] P. Cavanagh, G. P. Caplovitz, T. K. Lytchenko, M. R. Maechler, P. U. Tse, and D. L. Sheinberg, "The architecture of object-based attention.," *Psychonomic bulletin & review*, vol. 30, pp. 1643–1667, 4 2023.

[52] J.-M. Pierson, P. Stolf, H. Sun, and H. Casanova, "Milp formulations for spatio-temporal thermal-aware scheduling in cloud and hpc datacenters," *Cluster Computing*, vol. 23, pp. 421–439, 4 2019.

[53] M. Kansara, "A structured lifecycle approach to large-scale cloud database migration: Challenges and strategies for an optimal transition," *Applied Research in Artificial Intelligence and Cloud Computing*, vol. 5, no. 1, pp. 237–261, 2022.

[54] N. Aakash, B. Zhao, H. Zhu, J. Buryanek, J. Ding, Q. Sun, S. Zhang, K. Albin, M. Lerwill, M. Kem, R. H. Young, M. Mino-Kenudson, E. Oliva, S. Alghamdi, A. Pinto, K. Algashaamy, H. Alnajar, A.-L. Clarke, and W. Watkin, "Abstracts from uscap 2019: Gynecologic and obstetric pathology (993-1161).," *Laboratory investigation; a journal of technical methods and pathology*, vol. 99, no. Suppl 1, pp. 1–140, 2019.

[55] E. Winter, M. Rademacher, K. Shimotakahara, A. Surmann, R. Kohrs, M. Erol-Kantarci, K. Hinzer, M. S. Candidate, T. Riedel, D. Fellner, and J. Frederick, "Abstracts from the 9th dach+ conference on energy informatics," *Energy Informatics*, vol. 3, 10 2020.

[56] R. Rawassizadeh, T. Sen, S. J. Kim, C. Meurisch, H. Keshavarz, M. Mühlhäuser, and M. J. Pazzani, "Manifestation of virtual assistants and robots into daily life: Vision and challenges," *CCF Transactions on Pervasive Computing and Interaction*, vol. 1, pp. 163–174, 10 2019.

[57] E. W. Edobor and M. I. Marshall, "Earth, wind, water, fire and man: How disasters impact firm births in the usa," *Natural Hazards*, vol. 107, pp. 395–421, 2 2021.

[58] F. Steinbacher, W. Dobler, W. Benger, R. Baran, M. Niederwieser, and W. Leimer, "Integrated full-waveform analysis and classification approaches for topo-bathymetric data processing and visualization in hydrovish," *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 89, pp. 159–175, 5 2021.

[59] R. Halloush and H. Liu, "Modeling and performance evaluation of jamming-tolerant wireless systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 4361–4375, 11 2018.

[60] S. Gehr, N. K. Balasubramaniam, and C. Russmann, "Use of mobile diagnostics and digital clinical trials in cardiology.," *Nature medicine*, vol. 29, pp. 781–784, 3 2023.

[61] A. L. Carroll, S. C. Sillett, and R. V. Pelt, "Tree-ring indicators of fire in two old-growth coast redwood forests," *Fire Ecology*, vol. 14, pp. 85–105, 2 2018.

[62] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on tf-idf and lda schemes," *Human-centric Computing and Information Sciences*, vol. 9, pp. 1–21, 8 2019.

[63] W. Dekens, J. de Vries, and T. Tong, "Sterile neutrinos with non-standard interactions in - and 0-decay experiments," *Journal of High Energy Physics*, vol. 2021, pp. 1–29, 8 2021.

[64] A. K. Saxena, R. R. Dixit, and A. Aman-Ullah, "An lstm neural network approach to resource allocation in hospital management systems," *International Journal of Applied Health Care Analytics*, vol. 7, no. 2, pp. 1–12, 2022.

[65] H. Min, "Developing a smart port architecture and essential elements in the era of industry 4.0," *Maritime Economics & Logistics*, vol. 24, pp. 189–207, 2 2022.

[66] H. Beiki, A. L. Eveland, and C. K. Tuggle, "Recent advances in plant and animal genomics are taking agriculture to new heights.," *Genome biology*, vol. 19, pp. 48–48, 4 2018.

[67] Y. Yang, "Balanced scheduling method for big data of network traffic based on set-pair analysis strategy," *Journal of Physics: Conference Series*, vol. 2187, pp. 12065–012065, 2 2022.

[68] S. K. Sahay, N. Goel, M. Jadliwala, and S. Upadhyaya, "Advances in secure knowledge management in the artificial intelligence era," *Information Systems Frontiers*, vol. 23, pp. 807–810, 7 2021.

[69] E. Aamari and A. Knop, "Adversarial manifold estimation," *Foundations of Computational Mathematics*, vol. 24, pp. 1–97, 10 2022.

[70] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Integrating polystore rdbms with common in-memory data," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5762–5764, IEEE, 2020.

[71] R. Ahmad, I. Alsmadi, W. Alhamdani, and L. Tawalbeh, "Zero-day attack detection: a systematic literature review," *Artificial Intelligence Review*, vol. 56, pp. 10733–10811, 2 2023.

[72] Y. You, Y. He, S. Rajbhandari, W. Wang, C.-J. Hsieh, K. Keutzer, and J. Demmel, "Fast lstm by dynamic decomposition on cloud and distributed systems," *Knowledge and Information Systems*, vol. 62, pp. 4169–4197, 7 2020.

[73] J. M. Flores, G. Bourdin, A. B. Kostinski, O. Altaratz, G. Dagan, F. Lombard, N. Haëntjens, E. Boss, M. B. Sullivan, G. Gorsky, N. Lang-Yona, M. Trainic, S. Romac, C. R. Voolstra, Y. Rudich, A. Vardi, and I. Koren, "Diel cycle of sea spray aerosol concentration.," *Nature communications*, vol. 12, pp. 1–12, 9 2021.

[74] M. Al-Rakhami, A. Gumaei, M. A. Alsahli, M. M. Hassan, A. Alamri, A. Guerrieri, and G. Fortino, "A lightweight and cost effective edge intelligence architecture based on containerization technology," *World Wide Web*, vol. 23, pp. 1341–1360, 5 2019.

[75] Y. Lin, "Optimization and use of cloud computing in big data science," *Computing, Performance and Communication Systems*, vol. 7, no. 1, 2023.

[76] X. Gong, R. Jiao, A. Jariwala, and B. Morkos, "Crowdsourced manufacturing cyber platform and intelligent cognitive assistants for delivery of manufacturing as a service: fundamental issues and outlook," *The International Journal of Advanced Manufacturing Technology*, vol. 117, pp. 1997–2007, 8 2021.

[77] K. O. Said, M. Onifade, P. Akinseye, P. Kolapo, and J. Abdulsalam, "A review of geospatial technology-based applications in mineral exploration," *GeoJournal*, vol. 88, pp. 2889–2911, 11 2022.

[78] Y. Dong and Y.-D. Yao, "Iot platform for covid-19 prevention and control: A survey," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 49929–49941, 3 2021.

[79] S. Hedayati, N. Maleki, T. Olsson, F. Ahlgren, M. Seyednezhad, and K. Berahmand, "Mapreduce scheduling algorithms in hadoop: a systematic study," *Journal of Cloud Computing*, vol. 12, 10 2023.

[80] N. Barman, A. Borgohain, S. S. Kundu, B. Saha, R. Roy, R. Solanki, N. V. P. K. Kumar, and P. L. N. Raju, "Impact of atmospheric conditions in surface–air exchange of energy in a topographically complex terrain over umiam," *Meteorology and Atmospheric Physics*, vol. 131, pp. 1739–1752, 4 2019.

[81] M. Kesarwani, A. Kaul, G. Singh, P. M. Deshpande, and J. R. Haritsa, "Collusion-resistant processing of sql range predicates," *Data Science and Engineering*, vol. 3, pp. 323–340, 11 2018.

[82] W. Wang, J. V. Dinh, K. S. Jones, S. Upadhyay, and J. Yang, "Corporate diversity statements and employees' online dei ratings: An unsupervised machine-learning text-mining analysis," *Journal of Business and Psychology*, vol. 38, pp. 45–61, 5 2022.

[83] N. Berkani, L. Bellatreche, S. Khouri, and C. Ordonez, "The contribution of linked open data to augment a traditional data warehouse," *Journal of Intelligent Information Systems*, vol. 55, pp. 397–421, 2 2020.

[84] M. Kansara, "A framework for automation of cloud migrations for efficiency, scalability, and robust security across diverse infrastructures," *Quarterly Journal of Emerging Technologies and Innovations*, vol. 8, no. 2, pp. 173–189, 2023.

[85] C.-T. Yang, Y.-A. Chen, Y.-W. Chan, C.-L. Lee, Y.-T. Tsan, W.-C. Chan, and P.-Y. Liu, "Influenza-like illness prediction using a long short-term memory deep learning model with multiple open data sources," *The Journal of Supercomputing*, vol. 76, pp. 9303–9329, 2 2020.

[86] G. Ambrose-Igho, W. M. Seyoum, W. Perry, and C. M. O'Reilly, "Spatiotemporal analysis of water quality indicators in small lakes using sentinel-2 satellite data: Lake bloomington and evergreen lake, central illinois, usa," *Environmental Processes*, vol. 8, pp. 637–660, 4 2021.

[87] R. Rousseau, V. A. Glezakou, and A. Selloni, "Theoretical insights into the surface physics and chemistry of redox-active oxides," *Nature Reviews Materials*, vol. 5, pp. 460–475, 5 2020.

[88] A. Samanta and T. G. Nguyen, "Quality-driven energy-efficient big data aggregation in wbans," *IEEE Sensors Letters*, vol. 6, no. 8, pp. 1–4, 2022.

[89] L. Neely, S. Oyama, Q. Chen, A. Qutub, and C. Chen, "Tutorial: Lessons learned for behavior analysts from data scientists.," *Perspectives on behavior science*, vol. 47, pp. 203–223, 5 2023.

[90] A. Young and P. Rogers, "A review of digital transformation in mining," *Mining, Metallurgy & Exploration*, vol. 36, pp. 683–699, 7 2019.

[91] G. Smits, O. Pivert, R. R. Yager, and P. Nerzic, "A soft computing approach to big data summarization," *Fuzzy Sets and Systems*, vol. 348, pp. 4–20, 2018.

[92] L. Shi and Z. Wang, "Computational strategies for scalable genomics analysis.," *Genes*, vol. 10, pp. 1017–1017, 12 2019.

[93] R. B. Towbin, "Spr 2023," *Pediatric Radiology*, vol. 53, pp. 1–137, 5 2023.

[94] W. Bhimji, D. Carder, E. Dart, J. Duarte, I. Fisk, R. Gardner, C. Guok, B. Jayatilaka, T. Lehman, M. Lin, C. Maltzahn, S. McKee, M. S. Neubauer, O. Rind, O. Shadura, N. V. Tran, P. van Gemmeren, G. Watts, B. A. Weaver, and F. Würthwein, "Snowmass 2021 computational frontier compf4 topical group report storage and processing resource access," *Computing and Software for Big Science*, vol. 7, 4 2023.

[95] M. Shahin, F. F. Chen, A. Hosseinzadeh, and N. Zand, "Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: an early failure detection diagnostic service," *The International Journal of Advanced Manufacturing Technology*, vol. 128, pp. 3857–3883, 8 2023.

[96] D. C. Phan, T. H. Trung, V. T. Truong, T. Sasagawa, T. P. T. Vu, D. T. Bui, M. Hayashi, T. Tadono, and K. N. Nasahara, "First comprehensive quantification of annual land use/cover from 1990 to 2020 across mainland vietnam," *Scientific reports*, vol. 11, pp. 9979–9979, 5 2021.

[97] A. Korichi and T. Lauritsen, "Tracking γ rays in highly segmented hpge detectors: A review of agata and gretina," *The European Physical Journal A*, vol. 55, pp. 121–, 7 2019.

[98] S. L. Voelker, S.-Y. Wang, T. E. Dawson, J. S. Roden, C. J. Still, F. J. Longstaffe, and A. Ayalon, "Tree-ring isotopes adjacent to lake superior reveal cold winter anomalies for the great lakes region of north america.," *Scientific reports*, vol. 9, pp. 1–10, 3 2019.

[99] K. Crawford and T. Paglen, "Excavating ai: the politics of images in machine learning training sets," *AI & SOCIETY*, pp. 1–12, 6 2021.

[100] O. A. Obarein and C. C. Lee, "Differential signal of change among multiple components of west african rainfall," *Theoretical and Applied Climatology*, vol. 149, pp. 379–399, 4 2022.