المملكة العربية السعودية
وزارة التعليم
جامعة طيبة

جامعة طيبة
TAIBAH UNIVERSITY

Kingdom of Saudi Arabia
Ministry of Education
TAIBAH UNIVERSITY

Securing Cloud-Based Database Systems: Challenges and Best Practices in the Era of Big Data

(Advanced Database Management System)

By

Khadija S. Yahfath

4725302

Supervised by:

Dr. Tariq S. Mian

First Semester of the Academic Year 2025-2026

Abstract

The common use of Big Data and clouds raise serious challenges for securing sensitive databases. If you neglect these weaknesses, it can lead to significant damage, loss of reputation, and difficulty in complying with regulations.

The document assesses major threats to cloud-based database systems including data breaches, unauthorized access, and widespread misconfigurations. It then describes the complete architecture of up-to-date security solutions.

To enhance efficiency, businesses can use AI-enabled anomaly detection to reduce response time and utilize autonomous databases to automate administrative tasks with machine learning.

Also, to continuously remediate vulnerabilities, the framework uses Cloud Security Posture Management (CSPM) and foundational access controls such as the Principle of Least Privilege (PoLP) through advanced Identity and Access Management (IAM) and Multi-Factor Authentication (MFA).

Based on the study's conclusions, databases that operate in the cloud need to become adaptive and intelligent, and changes must also be compliance-driven to be transparent and fortify security.

Keywords: Cloud Database Security, Big Data, AI-driven Detection, CSPM, Identity and Access Management (IAM), Misconfigurations, Least Privilege, Regulatory Compliance

Table of Content

Contents

Table of Figures

1. Introduction

In today's digital landscape, cloud-based database systems have become a fundamental component of modern organizational data infrastructures. They play a vital role in processing the massive volumes of data generated daily in the era of Big Data. Operating within cloud computing environments, these databases provide superior flexibility, scalability, and accessibility compared to traditional on-premises solutions.

Leading providers such as Amazon Web Services (AWS), Google Cloud Platform, Microsoft Azure, and Oracle have developed advanced cloud database technologies that cater to both startups and large enterprises (Nadim, 2021).

The explosion of Big Data, driven by smartphones, IoT devices, social media, and AI workloads, has created an ever-growing demand for systems with high availability, resiliency, and scalability. Cloud-based databases address these demands by enabling dynamic resource allocation, seamless integration with machine learning tools, disaster recovery, and reduced management overhead. Moreover, these smart databases increase efficiency because they automatically manage routine administrative work.

However, this increasing reliance on cloud-based databases introduces several security challenges. Organizations now face threats such as data breaches, unauthorized access, insider attacks, and compliance risks, all of which can compromise sensitive information and disrupt business operations (Almubarak, 2021).

Therefore, this paper argues that implementing comprehensive security best practices is essential for protecting sensitive data, ensuring regulatory compliance, and maintaining both accessibility and scalability in cloud-based database systems within the Big Data era

2. Security Challenges

Cloud-based database systems face substantial security challenges due to their distributed nature and the vast amount of sensitive data they handle. Cloud-based database systems are responsible for storing sensitive data.

As businesses using cloud infrastructure shared by multiple tenants. Different types of vulnerabilities arise which are not common in On-Premises Environments. It's important to address these risks to guarantee data privacy, regulatory compliance, and smooth operations in the Big Data age.

2.1 Data Breaches

Among the most serious threats are breaches where an unauthorized entity can access your sensitive data. The multi-tenancy cloud database architecture increases risk since when data is breached, data belonging to many clients can be taken (Goel, 2024).

In May 2021, for example, a billion records of Cognyte were exposed due to weak authentication settings on the cloud database that contained data on earlier breaches (al., 2025).

Likewise, Microsoft scientists left as much of 38 terabytes of vital AI training data lolling unprotected in late 2023 because of errors with the setup of access privileges (Sidorkin, 2025).

Such incidents highlight the need for strong encryption, strong authentication and access management controls to prevent widespread data exposure in the cloud

2.2 Unauthorized Access

Unauthorized access refers to cases of unauthorized access of the database by attackers or malicious insiders who gain access to cloud databases. It can happen because of compromised credentials, malicious attacks or software exploit.

These attackers often use phishing, brute force, and/or IAM misconfiguration exploitation to take over accounts and evade systems. When they gain access, they can read, change, or delete sensitive data that disrupts business continuity and violate privacy policies.

For example, there has been an increase in the cloud account hijacking incidents whereby attackers take control of real accounts and either ransom or misuse the data. Exposed or weak API configurations also make it easy to access a service. In the context of Big Data, such incidents disrupt analytics pipelines and compromise data reliability (Kumar, 2023).

Enforcing multi-factor authentication and role-based access control remains essential for maintaining data confidentiality and integrity.

2.3 data Integrity Issues

Data integrity is the maintenance of, and the practice of, data accuracy and consistency. Data integrity issues in cloud databases result from malicious insiders, software bugs, application exploits (e.g., SQL injection), or ransomware attacks that can alter, corrupt, or delete data (al. W. B., 2022).

When data integrity is violated, it may lead to unacceptable analytics that can produce wrong decisions and bad operations. The situation is worse in the case of Big Data, where automation incorporates data without manual intervention and relies on error-free data streams.

For example, ransomware attacks or cybercrimes that encrypt the most important files in a database can make the data life-cycle work and cause the company a fortune in losses (Beaman, 2021).

To stop unauthorized changes to data, it is important to enforce rules such as auditing and real-time monitoring in a cloud security framework

2.4 Regulatory Compliance

Regulatory Compliance refers to laws and regulations governing data security and privacy. For example: GDPR, PCI-DSS and HIPAA. Meeting the requirements of various regulations on the cloud is a difficult task because of aspects such as multi-tenancy data architecture, data sovereignty that concerns localization and/or residency rules on data, and shared cloud service provider responsibility against security issues of the cloud. If you don't comply, you may be fined a lot and lose customers (Priyanshi Goswami, 2025).

For instance, in 2019, British Airways broke GDPR rules and faced a fine of £20 million from the UK Information Commissioner's Office (ICO) for the breach of personal data of over 400,000 customers. Research shows that having strategies in place to comply with regulatory standards and conduct during the storage of a cloud database is important (Bauskar, 2024).

2.5 Additional Challenges: Misconfigurations and Vulnerable APIs

Misconfigurations and insecure APIs represent another critical security challenge for cloud-based database systems. Cloud databases are vulnerable to threats due to mistakes in configuration and unprotected APIs in addition to the fundamental difficulties (Goel, 2024).

Databases may unintentionally become accessible to anyone on the internet due to incorrectly configured storage buckets, insufficient firewall restrictions, or excessively lenient network configurations. According to studies, misconfigurations account for roughly one-third of occurrences and are the primary source of cloud breaches (Bauskar, 2024).

Additionally, unprotected APIs, which serve as the entry points for database integration and management, might be used for injection attacks or illegal data access. To reduce these risks, it is essential to implement automated configuration management, enforce the least privilege principle, and conduct regular security audits (Fiser, 2024).

All these security challenges validate the paper's thesis that cloud-based database systems need a multi-layered approach based on best practices, security tools and compliance mechanisms. To be reliable, scalable and accessible, organizations need Data Management and Governance solutions that tackle the existing Data Breaches, Unauthorized Access, Data Integrity threats, Regulatory and Compliance Challenges, Configurations being Management vulnerabilities.

3. Best Practices

Securing cloud-based database systems requires comprehensive adoption of best practices designed to protect sensitive data, control access, and ensure compliance, all while maintaining the performance and scalability needed for Big Data applications.

This section outlines key strategies, including encryption, access control, regular security audits, and the deployment of advanced security tools, as essential components for robust cloud database security

3.1 Encryption

Encryption is foundational for protecting data both at rest and in transit within cloud database environments. Data encryption takes data and transforms it into ciphertext, which denies intruders or hackers' access to any meaningful information even if other defenses are breached (Shaik, 2023).

There are two main types of encryptions commonly used:

3.1.1 Symmetric encryption

in which data is encrypted and decrypted using a single secret key. Because they strike a mix between robust security and computing performance, algorithms like Advanced Encryption Standard (AES) are widely used (Kumari, 2022).

Transparent Data Encryption (TDE), which is extensively supported by cloud database providers such as Azure SQL and TencentDB, securely protects data "at rest" by encrypting disk-based data without necessitating modifications to application code (Madyatmadja, 2021).

3.1.2 Asymmetric encryption

which encrypts and decrypts data using a pair of related keys (public and private). This adds an additional layer to access management and is helpful for safe key exchange and safeguarding smaller data segments (Keswani, 2025).

Moreover, the use Transport Layer Security (TLS/SSL) protocols to encrypt data “in transit’’ between clients and the cloud server can prevent interception and tampering. Customer-Managed Encryption Keys (CMEK) controls key lifecycle management which is essential to comply with data protection regulations. Organizations manage these encryption keys (Steffen Muller, 2014).

3.2 Access Control

Access Control is a critical component of any robust security strategy, forming the fundamental bedrock of protection for cloud-based database systems.

This mechanism is designed to strictly manage and restrict all user and service interactions with sensitive data, directly countering the major threat of unauthorized access by ensuring that only verified entities can perform specific, authorized operations (M.S. Rahaman, 2023).

3.2.1 Role-Based Access Control (RBAC)

Effective access control systems limit who can access or change data in cloud databases. Role-Based Access Control (RBAC)) systems minimize superfluous privileges by allocating permissions according to individuals' roles and responsibilities (Akuthota, 2025).

3.2.2 Multi-Factor Authentication (MFA)

By demanding additional verification in addition to passwords, Multi-Factor Authentication (MFA) improves login security and stops unwanted access using credentials that have been stolen (Tran-Truong, 2025).

3.2.3 Identity and Access Management (IAM)

Solutions for Identity and Access Management (IAM) offer centralized control over user identities, policy enforcement, and real-time access attempt monitoring to identify irregularities early. Together, these defenses increase database secrecy and decrease the attack surface (Ezinwanneamaka, 2025).

3.3 Regular Security Audits and Monitoring

Continuous security audits examine configurations, permissions, and log data to identify vulnerabilities or breaches. Frequent evaluation guarantees that security measures continue to be effective in the face of changing threats.

Monitoring tools analyze system behavior and access patterns to flag unusual activities, such as unauthorized configuration changes or unusual data searches, and send out notifications for quick incident response. Maintaining compliance with laws like GDPR and HIPAA is another benefit of automated compliance inspections.

Building a culture of proactive auditing and continuous monitoring enables organizations to identify risks before they escalate, ensuring that data protection measures evolve alongside emerging threats.

3.4 Advanced Security Tools and Practices

Modern cloud environments increasingly rely on advanced security tools and practices that enhance visibility, automate threat detection, and minimize data exposure.

These tools such as Cloud-Native Application Protection Platforms (CNAPPs), Intrusion Detection and Prevention Systems (IDS/IPS), and Data Loss Prevention (DLP) with anomaly detection form an integrated defense layer that safeguards cloud-based database systems from both external and internal threats (KuppingerCole, 2025).

3.4.1 Cloud-Native Application Protection Platforms (CNAPPs)

Cloud-Native Application Protection Platforms (CNAPPs) offer unified protection for workloads, APIs, and configurations across multi-cloud environments. These platforms provide end-to-end visibility into potential threats and automate compliance monitoring to maintain a strong security posture.

By integrating multiple capabilities, including vulnerability management and configuration analysis, CNAPPs help organizations manage complex environments more effectively.

3.4.2 Intrusion Detection/Prevention Systems (IDS/IPS)

Intrusion Detection and Prevention Systems (IDS/IPS) play a vital role in identifying and blocking malicious activity before it affects cloud resources (Nasim', 2025).

The 2019 Capital One data breach, where sensitive client records stored on AWS were exposed due to a misconfigured web application firewall, underscores the importance of continuous monitoring and automation to maintain configuration integrity (Capital One, 2019). These systems, when properly configured, can detect such vulnerabilities early and prevent large-scale data exposure (Neto, 2020).

3.4.3 Data Loss Prevention (DLP) and Anomaly Detection

Data Loss Prevention (DLP) tools, combined with machine learning–based anomaly detection, enable organizations to recognize subtle irregularities that may indicate security breaches. Anomaly detection excels at identifying unusual behavior patterns that deviate from normal operations (Abid, 2024).

Furthermore, data masking techniques enhance the safety of analytics, training, and testing processes by concealing sensitive data fields without exposing real information. This dual approach strengthens confidentiality while allowing data-driven innovation.

In conclusion, protecting cloud-based database systems requires a multi-pronged defense approach that includes encryption, strict access control, continuous audits, and the use of advanced cloud security solutions.

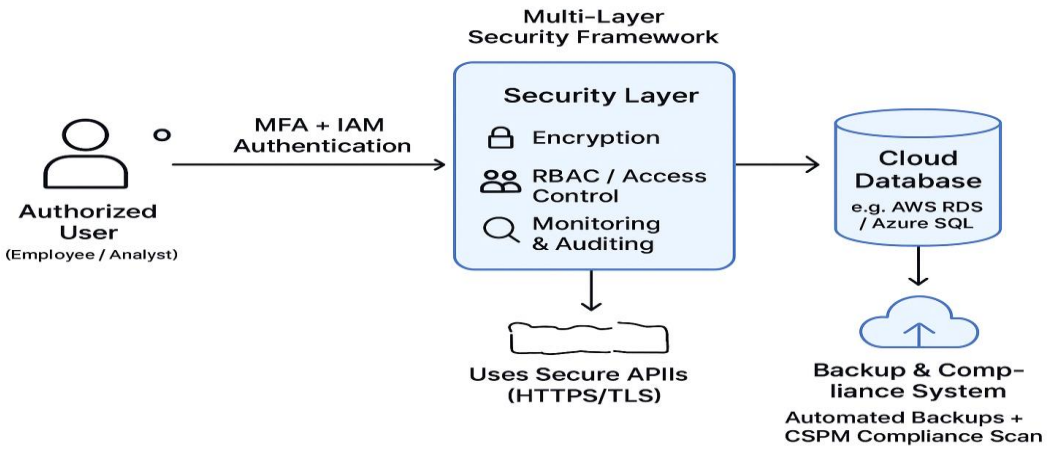

Figure 1. Secure Cloud Database Architecture

The figure illustrates the proposed multi-layered security framework for protecting cloud-based database systems, showing user authentication, encryption, access control, monitoring, and backup layers

Together, these best practices protect private information, maintain regulatory compliance, and maintain the scalability and accessibility required for efficient Big Data management in cloud environments, all of which directly support the main idea of this study.

4. Case Studies of Cloud Database Security Incidents

Cloud database security must be understood through the viewpoint of actual incidents. The theory on security problems and best practices presented in earlier sections is helpful. However, it is not enough by itself.

This section will look at some actual instances of security failures instead of studying theoretical vulnerabilities.

4.1 Football Australia AWS S3 Bucket Misconfiguration (2024)

One notable example occurred in 2024 when Football Australia experienced a data breach due to misconfigured AWS S3 buckets, which exposed personal information of players and staff. The issue stemmed from developers embedding plaintext AWS access keys in website source code and leaving storage buckets publicly accessible without proper access controls.

Following the breach, Football Australia restricted access to the exposed buckets, revoked compromised API keys, and conducted an organization-wide security audit to detect similar vulnerabilities. They also implemented stricter Identity and Access Management (IAM) policies and provided secure coding training to developers to prevent embedding credentials in source code (CSA, 2025).

This response demonstrated effective incident management; however, the root cause combination of insecure default configurations and weak development practices exposed systemic vulnerabilities.

The case highlights the critical importance of continuous developer training, automated configuration validation, and proactive governance within the Software Development Lifecycle (SDLC).

4.2 Major Cloud Storage Provider Data Breach (2024)

in mid-2024, a major cloud storage provider suffered a significant breach, which revealed sensitive business data from various enterprise clients. The hackers attacked because the company was not Multi-Factor Authentication (MFA) which allowed them to steal credentials.

In response, the provider made MFA mandatory, enhanced authentication workflows with conditional access policies, and launched awareness campaigns, and launched awareness campaigns to enhance user credential hygiene along with phishing attacks (Team, 2025).

These corrective measures strengthened account security and prevented similar incidents, though the breach initially revealed a major gap in enforcing access control policies. Cloud infrastructures can be compromised by poor authentication practices. It's essential to enforce strong security defaults and educate users continuously. This case shows the value of these practices.

4.3 Facebook Third-Party Developer Data Leak (2019)

Another illustrative case occurred in 2019 when third-party developers exposed over 540 million Facebook user records stored on Amazon AWS due to cloud misconfigurations. The exposed data lacked encryption and proper monitoring, remaining publicly accessible for months, although no passwords were compromised.

Following the incident, Facebook collaborated with AWS to restrict public data access, enforced stricter third-party data governance policies, and deployed automated scanning tools to detect insecure configurations in the future (Moss, 2019).

The company's response effectively contained the incident; however, it underscored the risks of third-party integrations and insufficient monitoring controls. This case highlights the

necessity of robust third-party risk management frameworks and automated compliance checks to ensure cloud systems maintain confidentiality and integrity.

In conclusion, cloud-based database systems provide flexibility, scalability, and operational efficiency, yet they remain vulnerable to security issues such as data breaches, misconfigurations, and insider threats.

Organizations must adopt strong encryption, enforce strict access management, and conduct continuous monitoring and auditing to mitigate these risks.

By understanding real-world incidents and implementing proactive best practices, organizations can enhance data protection, maintain compliance, and ensure the reliability and resilience of their cloud database systems.

5. Balancing Security and Accessibility

Cloud databases face the challenge of balancing security measures with accessibility requirements, especially in Big Data environments, where scalability, rapid data availability, and performance are critical.

Stricter access controls or reduced system flexibility are sometimes necessary to secure cloud databases through encryption, monitoring, and identity management.

However, these controls can potentially hinder legitimate users from accessing data efficiently, which may affect business operations and decision-making.

5.1 Trade-offs Between Security and Accessibility

Implementing robust security measures, including Multi-Factor Authentication (MFA), Role-Based Access Control (RBAC), encryption, and continuous auditing, strengthens data protection but can increase system complexity and latency. For instance:

- Encryption may introduce computational overhead that can slow query response times.
- Strict access policies can delay authorized users from retrieving data promptly, particularly in distributed teams or high-volume analytics scenarios.

Conversely, prioritizing accessibility without sufficient controls can expose organizations to severe risks, such as unauthorized access, data leaks, or large-scale breaches.

The distributed and multi-tenant nature of cloud environments amplifies the impact of such vulnerabilities, making it essential to carefully balance usability and security (Akrer, 2024).

5.2 Strategies to Achieve Balance

Organizations can adopt several strategies to maintain this balance:

5.2.1 Classifying and Segmenting Data

Apply stricter controls to highly sensitive or regulated data while allowing greater flexibility for less critical datasets.

5.2.2 Adaptive Access Controls

Implement dynamic permissions based on user roles, device security status, location, and risk profiles. Identity and Access Management (IAM) solutions can enforce context-aware policies that maintain security without hindering usability.

5.2.3 Layered Encryption

Combine encryption at rest with selective field-level encryption to protect critical data efficiently without excessive performance costs.

5.2.4 Continuous Monitoring and Automation

Deploy automated tools to monitor access patterns, detect anomalies, and respond quickly to threats without disrupting normal workflows.

5.2.5 Hybrid and Multi-Cloud Architectures

Store sensitive workloads in private or on-premises environments while using public clouds for less critical operations. This approach requires strong governance, IAM, and compliance frameworks to ensure security across environments.

5.3 Insights for Organizations

To effectively balance security and accessibility, organizations should adopt a risk-based approach that includes:

- Understanding business priorities and regulatory requirements to tailor security measures.
- Engaging cross-functional teams (security, IT, compliance, and business units) to ensure policies align with operational realities.
- Investing in scalable cloud security solutions that integrate seamlessly with existing workflows and support automation.

- Regularly reviewing and updating access policies and configurations as threats evolve and business needs change.

6. Conclusion

Ultimately, organizations that integrate these best practices and proactive governance measures can achieve robust, scalable, and reliable cloud database environments. Future research should explore AI-driven security solutions, hybrid cloud strategies, and enhanced regulatory compliance automation to further strengthen cloud database security in the era of Big Data.

With Big Data systems growing more complex, advanced in speed and larger in scale, cloud database security has become necessary to ensure confidentiality, availability and integrity of such systems. As cloud computing has grown in use, organizations are storing and processing more information in the 'Cloud'.

As one can imagine, this has increased the exposure to sophisticated cyber threats. It is essential that these systems are ensured with sound and fitting protection that includes a mixture of policies, practices and technologies.

A multi-layered security framework to protect cloud databases must include the following features and techniques:

- AI-Powered Threat Detection and Automated Response**:**

AI and machine learning models analyze extensive data in the cloud to detect any anomaly like un-authorized access attempts or data theft. The app does this analysis in real-time. AI-driven systems lessen reliance on static rules. They adopt new patterns of attack. Response times before breach are cut down.

- Cloud Security Posture Management (CSPM):

CSPM tools will automatically monitor your cloud infrastructure for misconfigurations and remediate any issue before they lead to cloud breach. They scan cloud assets frequently, alerting security teams to any vulnerabilities found as well as deviations from security policy standards.

- Identity and Access Management (IAM):

Using the least privilege principle reduces unauthorized access points through role-based access controls and multifactor authentication.

- Regulatory Compliance and Data Governance:

A critical element of cloud database protection is ensuring compliance with data protection regulations and industry standards. Incorporating compliance checks and automated auditing within the security framework assists organizations in meeting regulatory requirements as well as ensuring transparency and accountability.

These security measures like the standard control at the user side provide additional security to cloud databases while not interfering with Big Data tasks. Organizations must increasingly deploy adaptive, intelligent, compliance-driven technologies to safeguard their cloud environments

References

Akrer, S. S. A. (2024). Balancing Data Accessibility and Security in Cloud-Based Business Intelligence Systems. ***American Journal of Technology Advancement***.

Almubarak, A. (2021). Cloud database security challenges and solutions in the big data era. ***Journal of Physics: Conference Series, 1879***(2). https://doi.org/10.1088/1742-6596/1879/2/022026

al., W. B. (2022). **Big data security: Challenges and solutions** (2nd ed.). IGI Global.

Akuthota, K. K. S. (2025). **Exploring cloud data security strategies: A review of IAM and encryption**. IEEE.

Bauskar, R. (2024). **Cloud security posture management: A guide to preventing misconfigurations and breaches**. O'Reilly Media.

Beaman, M. (2021, July 1). *Ransomware attacks surge: Why data integrity matters more than ever*. **Data Protection Journal**. Retrieved from: https://www.dataprotectionjournal.com/ransomware-attacks-surge-why-data-integrity-matters-more-than-ever/

Capital One. (2019, July 29). *Capital One announces information security incident*. **Press Release**. Retrieved from: https://www.capitalone.com/press/capital-one-announces-information-security-incident/

CSA. (2025). *Cloud misconfigurations and the supply chain: Lessons from Football Australia*. **Cloud Security Alliance Blog**.

Ezinwanneamaka, C. (2025). **Advanced identity and access management for multi-cloud environments**. Springer.

Fiser, T. (2024, February 1). *Insecure APIs: The silent threat to cloud data*. **Cloud Security Today**.

Goel, P. (2024). *Multi-tenancy and data breach risk in cloud databases*. **Journal of Cloud Computing, 13**(1), 1–15. https://doi.org/10.1186/s13673-024-00123-y

Keswani, V. (2025). **Asymmetric encryption: Security and key management in modern systems**.

Kumari, P. (2022). *Symmetric encryption algorithms: A comparative analysis*. **International Journal of Information Security, 21**(4), 650–670. https://doi.org/10.1007/s10207-021-00567-z

Kumar, S. (2023). **The impact of unauthorized access on big data analytics pipelines**. Taylor & Francis.

KuppingerCole. (2025). **CNAPPs vs. traditional security: A comprehensive guide**. KuppingerCole Analysts.

Madyatmadja, G. S. (2021). **Transparent data encryption (TDE) for cloud database security: Implementation and performance**. IOP Conference Series: Earth and Environmental Science, 866(1). https://doi.org/10.1088/1755-1315/866/1/012001

M.S. Rahaman, D. (2023). **Access control mechanisms in cloud computing: A comparative study**.

Moss, S. (2019, April 3). *Facebook data leak exposes 540 million records on Amazon cloud*. **Tech Security News**. Retrieved from: https://www.techsecuritynews.com/facebook-data-leak-exposes-540-million-records/

Nadim, M. (2021). **Cloud database platforms: AWS, Azure, Google, and Oracle compared**. Packt Publishing.

Nasim', A. (2025). **Intrusion detection and prevention systems in cloud environments**.

Neto, V. (2020). **Web application firewall (WAF) misconfiguration: Lessons from the Capital One breach**.

Priyanshi Goswami, S. (2025). **Regulatory compliance in cloud computing: GDPR, HIPAA, and PCI-DSS requirements**.

Shaik, S. (2023). **Fundamentals of data encryption in cloud computing**.

Sidorkin, A. (2025). *The human factor: Microsoft AI data leak highlights configuration errors*. **Cybersecurity Quarterly**.

Steffen Muller, B. (2014). **Managing customer-managed encryption keys (CMEK) in the cloud**.

Team, T. R. (2025). *MFA bypass and credential theft: The 2024 cloud storage provider breach*. **Threat Research Blog**.

Tran-Truong, H. (2025). **Multi-factor authentication: Implementation and future trends**.

(al., 2025). (2025). **Cognyte data exposure: A case study in weak authentication**.