happiest minds
The Mindful IT Company

Data masking

Big Data Environment Best Practices

Introduction

Many organizations accidentally breach the Data Protection while moving sensitive information from Production to non-production environments, this puts customers' confidential information at stake and the possibility of data privacy breach is high when the data goes to the hands of inappropriate stakeholder. However, the necessity of production data is high since developers, and analytic researchers rely on the factual data for their strategy, research, and development. Moving the sensitive information from production to other environments will lead to noncompliance of data regulations like GDPR, PCI DSS, HIPAA, and so on. Therefore, there is a need for a tool that can remediate this gap and enables smooth transfer of information to the lower environment without revealing the sensitive information also comply with Data regulations.

Big Data

Big Data is a collection of large data sets with Structured, Semi-structured and unstructured data types that can be analyzed computationally to reveal patterns and trends from the stored data. Recently many organizations managing and utilizing bulk data sets such as social media, airlines, streaming, satellite, and IoT are deploying Big Data platforms to store sensitive and high-velocity data.

Big Data Security in Cloud Computing

Since the organizations have hugely adopted the Big data platforms, the security of these platforms, need to be tightened as it involves the usage and storage of sensitive information.

Big Data Security in Cloud Computing is vital due to the following issues:

- To protect and prevent Large size of sensitive business, defense, or regulatory information's from unauthorized user access
- To Prevent users from copying confidential information to the lower environments like UAT for research and analytics
- To minimize risk while sharing the data within the organization or with third parties
- To comply with data regulation standards like GDPR, PCI, PII, and PHI, and so on.

Best Practices to deal with Big Data Risks and Data masking

Understand the Data

Before masking any datasets, it is primal to understand the type of data that needs to be secured. Because it is impossible to secure every piece of data in a Big Data environment. Doing that will utilize the entire resources and affect the other data processing in Big data cloud. Therefore, before proceeding with the data protection solutions, evaluate the sensitive data that needs to be moved to a lower environment or which needs to be shared with the research team to understand the pattern. Prioritize only those data for protection, discover and identify highly sensitive information, such as personally identifiable information (PII), personal health information (PHI) and other intellectual property from the stored data in Big data cloud.

Defining Policies

Masking policy is essential to protect all sensitive data. As the data will be used for development, testing, backup, and analytics purposes it is mandatory to understand the available data and what it points at. Ideal data masking technique is selected based on sensitive data present and its location. While implementing one must consider system architecture, estimated data sizing, and business needs.

Data Masking Vs Encryption

There are some similarities between data masking and encryption, but in operational, both serve a different purpose. Data masking is one-way process once the data is masked it cannot be de-masked but in the case of Encryption techniques such as FPE (Format Preserved Encryption) the data can be encrypted and decrypted using the encryption keys. Therefore, the selection of techniques relies on the use case. Encryption helps protecting data from external attacks and data masking is for safeguarding sensitive data internally.

Data integrity

After the masking process is complete, ensure that the data integrity is maintained. Quality assurance and testing are performed to ensure masking configurations are producing the intended results.

Review the Masking Performance and Maintenance

It is paramount to assess masking performance and bring in necessary changes in masking configuration to optimize performance by fine-tuning the policies, sizing of the data and adopted masking techniques. Review reports and systems logs enable effective and efficient maintenance. Knowledge transfer, user training, and handoff ensure proper understanding of the selected masking process. Maintaining procedure documents and active change management acts as a knowledge base.

Conclusion

All organizations should ensure reliable protection for sensitive data to meet compliance requirements and prevent a data breach. Data breaches and loss of data can have a negative impact on the business, including legal and financial losses. Discovering sensitive data and administering appropriate masking and encryption techniques help organizations to meet regulatory requirements and customers' call to protect raw data.

About Happiest Minds Technologies:

Happiest Minds enables Digital Transformation for Enterprises and Technology providers by delivering seamless Customer Experience, Business Efficiency and Actionable Insights through an integrated set of Disruptive Technologies: Big Data Analytics, Internet of Things, SDN & NFV, Mobility, Cloud, Security, Unified Communications, etc. Happiest Minds offers domain centric solutions applying skills, IPs and functional expertise in IT Services, Product Engineering, Infrastructure Management and Security. These services have applicability across industry sectors such as Retail, Consumer Packaged Goods, Ecommerce, Banking, Insurance, Hi-tech, Engineering R&D, Manufacturing, Automotive and Travel/Transportation/Hospitality. Headquartered in Bangalore, India, Happiest Minds has operations in the US, UK, Singapore, Australia and has secured $63 million Series-A funding. Its investors are JPMorgan Private Equity Group, Intel Capital and Ashok Soota.

About Author

Sathish Kumar works with Happiest Minds Technologies as Technical Lead – Data Security Practice. He has an overall experience of 8.3 years in Data Security and Cyber Security. He has extensive experience in implementing various Data Security products in the Big Data environments, Databases, Exchange, Fileservers and so on.