Data Compliance Architecture for Managing Organizations Local and Global Data Using Data Virtualization

Mr. Rahul Shivaji Bagade[1], Dr R G Pawar[2] & Dr Anuradha Dandnaik[3]

[1]Research Scholar, Pune.
rahulbb99@gmail.com
[2]Director, Department of LifeLong Learning & Extension, Shivaji University, Kolhapur.
rgpawar@rediffmail.com
[3]Associate Professor, Neville Wadia Institute of Management Studies and Research, Pune.

Abstract—*In today's globalized world organizations crossing boundaries and doing organizations in various countries and regions. These organizations work in various sectors and industries. With strict data compliance and data privacy laws, organizations need to adhere local laws and store sensitive data in countries boundary and need to abide local land laws regarding data compliance. To sustain and compete in a competitive environment organizations make use of data they generate. The data consumption depends on the location of head office or location of top higher management. To consume data for various needs, organizations required data compliance and privacy protection data management framework.*

The aim of this paper is to study how a data virtualization platform can help organizations to adhere data compliance and data privacy. To maintaining balance between local and global data needs based on data sensitivity and further help to build data compliance system architecture using data virtualization solutions.

Keywords: *Data compliance, data localization, compliance framework, data virtualization, data compliance management system*

Objective

To study data compliance challenges faced by organizations

To study data compliance framework

To analyse role of data virtualization in mitigating compliance requirements

To generalized data compliance architectures

INTRODUCTION

In a globalized environment, organizations work in many countries and regions. These organizations work in various sectors such as finance, healthcare, pharma, banking, food safety, information technology, insurance, environment, intellectual property, security exchange, telecommunication etc. Each organizations generates vast amount of data based on size and sector they are working. The data includes organizations data, personal identification information, medical information, product data, service data and other sensitive information. Organization need both global and local data for their business requirements, oprtations and analysis. To protect the privacy and security of personel information many contries laid down regulations also know as data compliance.

Despite these regulation to thrive in todays competative world organizations need to use data for many purposes while adhering to regulations. To do so organization s need to manage local & global datasets based on data sensitivity.

ORGANIZATIONS NEED FOR LOCAL DATA AND GLOBAL DATA

Local data means generating, collecting, processing, storing, sharing and reporting sensitive compliant data within regulators' confined boundary on the other hand global data from data usage point means using data for various gains such as competitive advantage, sentiment analysis, advance healthcare, better product and service development.

Globally new data compliance regulations are promoting data localization. There are numerous reasons to do so but for organizations it poses challenges in numerous ways including growth, cross border organizations, data analysis, data movement for various gains such as encourage economic competitiveness, new product and service development, product and service review analysis, various organizations focus on analysis such as sales, marketing and man power (human resource) analysis. Managing compliance for local and global data requires a comprehensive approach that takes into account the specific regulations that apply to each jurisdiction where the organization operates.

Steps for managing compliance

1. Regulations - Identify applicable regulations
2. Establish policies and procedures
3. Implement technical and organizational measures
4. Conduct regular audits and reviews
5. Build centralized system

COMPLIANCE

Data compliance refers to the practice of adhering to laws, regulations, and standards related to the collection, processing, storage, and sharing of data. This includes protecting the privacy and security of personal information and ensuring that data is used in a legal and ethical manner. Data compliance is the process of achieving the minimum required data protection made mandatory by compliance authority in law. To be compliant means organizations has to take and implement measures, process, procedures, workflows, frameworks and operations layout designed to ensure the obligations made by authority are fulfilled.

In other teams it means the process of adhering to various regulatory standards, laws, governance, and legislations to maintain integrity, and availability of data. Data compliance include data source, data processing, data collection and data reporting of regulated data such as personal and sensitive data. Data privacy is the concept of how an individual's personal information is collected, processed, stored and shared. Data compliance comes in practice to help ensure the protection of regulated and/ or sensitive data from unauthorised access and use. Data compliance also tracks what kind of data and how much data is being stored and how this stored data is being utilized and managed throughout its data lifecycle.

Here are some of reason for data compliance guidelines:

- Data security (Data misuse and unauthorized access)
- Customer trust
- Rapid data growth
- Competitive advantage
- Use of data for market analysis, sentiment analysis
- Cyber security threats
- Mobile data such as social media content
- Data privacy
- Access to data by law enforcement agencies
- National security

Challenges posed by data compliance to organizations:

Data compliance pose many challenges to organizations, are as following:

- Complexity of regulations
- Constantly evolving regulations
- Risk of non-compliance
- Financial cost (Penalties)
- Legal risk
- Operational disruption
- Reputation damage
- Technical challenges
- Regulatory challenges
- Impact on business operations
- Lack of resources
- Technology constraints
- Data silos (Volume, Velocity, Variety)
- Human errors
- Protection of the vulnerable
- Conflict of laws

Around the globe there are 71% countries with data compliance legislations, 9% countries with draft legislation, 15% countries have no legislation. These countries follow several data compliance standards. Compliances such as GDPR, HIPAA, PIPEDA, CCPA, PCI DSS, FACTA, ISO/IEC, COPPA, POPIA, NIST, SOX, PDBP, DCIA, PIPL, FedRAMP etc. and organizations need to comply with these standards at any cost, so they can pursue their organizations depending on their operation and location in which these organizations operates.

Following measures organizations can take to handle data compliance and overcome challenges:

- Access and identity control
- Set up data compliance architect (including Protection against malware and cyber attacks)
- Build data flow pipelines (Including Data loss prevention, Control over data sharing)
- Set up alerting and notification system (Monitoring, Incident response and reporting)
- Acquire data compliance management system (Compliance Review)
- Develop organizations continuity plan (Disaster Recovery)

- Acquire auditing system (Internal & External Audit)
- Train & improve people for data security & privacy
- Regulatory compliance training for compliance obligations
- Make sure data protection measures are up to date
- Maintain records of data protection measures and audit procedures
- Apoint data security and compliance standards personnel (Compliance Officer)
- Adopt common control framework
- Establish and maintain your policies and procedures
- Continual improvement

Cross Regulatory Compliance Strategy

Cross regulatory compliance strategy determines in what ways data privacy regulations overlap in order to synergize compliance efforts. The most common requirements include protection of sensitive data and data protection impact assessments, sensitive data retention process, procedures and policies, data breach alert and notifications. To achieve data compliance one has to achieve the minimum standard outlined by the authorised body in data protection compliance law.

DATA COMPLIANCE MANAGEMENT SYSTEM

To address challenges arises due to compliance, businesses need to implement data compliance management system and establish clear policies and procedures for data collection, processing, sharing, and storing. Data compliance management system is framework that helps organizations manage their compliance with data protection regulations. The system includes policies, procedures, tools and technologies. The system helps organizations monitor, control and protect the personal data they collect, process and store.

DATA COMPLIANCE FRAMEWORK

A compliance framework is an important data management system for organizations to manage their compliance obligations and ensure that they are operating within legal and regulatory requirements. A data compliance framework is a set of policies, procedures, and practices designed to ensure that an organizations complies with legal, regulatory, and industry requirements related to data protection and privacy.

There are many reasons why organizations need a compliance framework, including:

- Legal requirements: Many organizations are subject to legal and regulatory requirements that require them to comply with specific rules and regulations.
- Risk management: Compliance frameworks are designed to help organizations manage risk by identifying potential areas of non-compliance and implementing controls to mitigate those risks.
- Efficiency: A compliance framework can help organizations to streamline their compliance activities by providing a structured approach to compliance management.
- Competitive advantage: A compliance framework can also provide organizations with a competitive advantage by demonstrating their commitment to compliance, risk management, and responsible organizations practices.
- Continuous improvement: A compliance framework provides a structured approach to compliance management that enables organizations to monitor their compliance performance, identify areas for improvement, and implement corrective actions.
- Centralized system: A compliance framework provides a centralized unified view of data across multiple data sources.

Benefits of implementing a data compliance framework include:

- Reduced risk of legal and regulatory penalties
- Increased customer trust and loyalty
- Improved data security
- Enhanced reputation and brand value
- Increased operational efficiency
- Reduced human errors
- Mitigate penalties
- Build a Positive Reputation among Employees, Customers, and the Public
- Achieve Higher Employee Productivity and Higher Employee Retention

CENTRALIZED SYSTEM

Data virtualization is useful tool for data compliance as it provides a unified view of data across multiple sources located round the globe in different regions and contries. Centralized system provides a unified solution for managing local and global data with comprehensive approach.

Benefit of using centralized system for data compliance:

- Data privacy
- Data governance (Centralized platform for data lineage, data quality, and data security policies)
- Data integration
- Data access
- Data management

DATA VIRTUALIZATION (LOGICAL DATA WAREHOUSE)

Data virtualization is a technology that allows data to be accessed and manipulated without physically moving or replicating it. It provides a unified view of data from multiple sources, including databases, data warehouses, and cloud-based applications, by creating a virtual layer over the physical data sources. Data virtualization help organizations manage data compliance by providing a unified view of data across multiple sources, implementing data privacy and data access controls, and centralizing data governance and data management policies and procedures.

Data virtualization enables organizations to improve analytics effectiveness and reduce analytics infra and manpower cost by building virtual data abstraction layers from diverse data sources. Data virtualization enables organizations to integrate, access and fetch data from multiple sources without moving data from one source to be integrated into target data store. Data virtualization is a data management platform, it provides consolidated and integrated data to consumers and downstream applications such as api's and analytical tools without data movement which is required in traditional data consolidation and integration approach. In nutshell data virtualization is a platform that allows organizations to access and integrate data from multiple sources without physically moving, duplicating or replicating the data.

Role of data virtualization in data compliance:

Data virtualization provides a unified and integrated view of data, which can help organizations to improve their data compliance practices in ways mentioned below:

- **Data governance:** Data virtualization can help organizations to improve their data governance by providing a centralized integrated view of data that is consistent, accurate, and up-to-date. This can help organizations to better manage data quality, access controls, and data lineage, which are important aspects of data compliance.
- **Data security:** Data virtualization can also help organizations to enhance their data security by providing a layer of abstraction between data sources and end users. This can help to reduce the risk of data breaches and unauthorized access to sensitive data, which are important compliance requirements.
- **Streamlined compliance reporting:** Data virtualization can help organizations to streamline compliance reporting by providing a centralized view of data that is consistent across different systems and applications. This can help to reduce the time and effort required to prepare compliance reports and ensure that the reports are accurate and up-to-date.
- **Improved data privacy:** Data virtualization can also help organizations to improve data privacy by providing a mechanism for controlling access to sensitive data. By providing a centralized view of data, organizations can more easily identify and manage data that is subject to privacy regulations.
- **Data masking:** Data virtualization can also be used to mask sensitive data by replacing it with a non-sensitive value. This can help organizations to comply with data privacy regulations by ensuring that sensitive data is not exposed to unauthorized users.
- **Data lineage:** Data virtualization provides a complete view of data lineage, allowing organizations to track data usage and ensure that data is being used in compliance with legal and regulatory requirements. This can help organizations to demonstrate compliance with data privacy regulations.
- **Data quality:** Data virtualization can also help organizations to ensure that data is accurate and consistent, which is an important requirement for compliance with regulations. By providing a real-time view of data from multiple sources, data virtualization can help organizations to identify data quality issues and take corrective action.
- **Compliance reporting:** Data virtualization provides a centralized view of data from multiple sources, making it easier for organizations to generate compliance reports and demonstrate compliance with legal and regulatory requirements. This can help to reduce the administrative burden of compliance reporting and ensure that organizations are meeting their compliance obligations.
- **Data protection:** Data virtualization can facilitate data protection by allowing organizations to apply data masking, anonymization, or encryption techniques at the virtualization layer. This can help to ensure that sensitive data is protected, even when it is being used in multiple applications or shared

with external partners.

- **Reduce compliance costs:** Data virtualization can help to reduce compliance costs by providing a more efficient and cost-effective alternative to traditional data integration approaches that require data to be physically moved or replicated. This can help organizations to manage their compliance obligations more efficiently, and can reduce the cost of compliance.
- **Data localization:** Data virtualization can help to build sensitive data localization within physical boundaries of regulatory jurisdiction. Which deeply helps organizations to localize critical, personal and sensitive data.
- **Data globalization:** Data virtualization can help to build non compliant data gateway for access from other locations. This data helps to build analytics to gain insights from underlying data.
- **Summarize data:** Data virtualization can help to build data summaries with help of virtual integration of localized summaries to total summaries of most critical key performance indicators of organizations. With data virtualization there is no need for data movement for creating summaries from multi-region, multi-country spanned data silos.
- **Data to insights turn around time:** Data virtualization greatly reduces the time required to generate insights from dispersed data silos.
- **Virtual data model reusability:** The data models prepared in a data virtualization platform can be reused or repurposed for different case scenarios.
- **Change management:** Data virtualization reduces the time required to implement changes which are a vital part of data compliance as there is no need for data movement.
- **Data pipelines:** The pipelines built within data virtualization platform helps compliance personnel about all the changes made to data with help of data lineage capability. It helps to back track issues if they arise due to not following data compliance processes.

Data virtualization layer handles:

- Real time data access
- Data integration without need of data replication
- Data transformation
- Data federation (collect, cobine and exchange data across multiple data sources)
- Logical data integration across all sources from different geographis
- Prompt in handling changing business requirements
- Improved data governance
- Data privacy
- Data security & protection
- Data marking
- Data lineage
- Data access control (Centralized)
- Helps in managing local and global datasets
- Helps in complying with data protection regulations

DATA VIRTUALIZATION BLOCK SOLUTIONS

In order to transform raw data to meaningful insights, organizations need to integrate data from multiple sources, the source data sets are actionable, trustful and raw in nature. To gain insight from raw data organizations need to infuse organizations logic, which enhances the value of data. Doing so traditional practices are not with time to handle various data related aspects, such as data compliance and data privacy.

To manage compliance related concerns, organizations need to adopt more advance techniques such as data virtualization techniques, which is industry proven practice followed by various organizations. Data virtulization tackel challenges pose by data complience by differentiating between local data sets and global data sets. Which further helps organizations to draw and segregate data based on compliance regulations and make use of data silos for building insights.

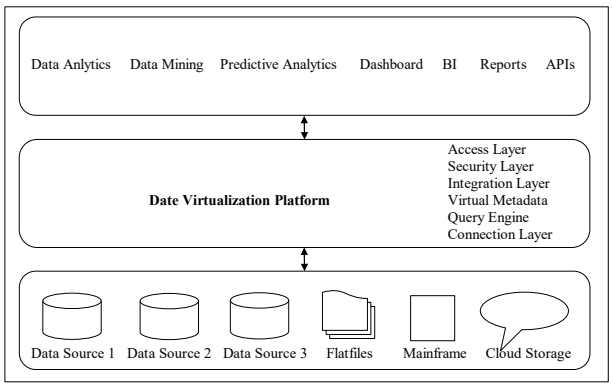

Image 1: Data Virtualization Bloack Architecture

The data virtualization block architecture generalised for more robust adaptation includes 3 major components, data sources, data virtualization platform and data consumption platforms respectively. The sources are of various kinds

and need to integrate data despite their physical presence due compliance norms.

In data virtualization layer, data from various data silos and systems are inegrated, abstracted and secured with various techniques. This give organizations an wholesome view of entire data lanscape. In this layer data is not moved physically, by using virtual master and worker nodes they are integrated as per requirement and as per use case. Worker nodes reside in specific geography and act as data center and abstraction layer, and based on compliance norm it feeds data to master node. Master node on other hand do the integration and security management. The settings of master node and worker node are not standard, as per requirement can be modified.

Consumption layer uses the underlaying platform for insight development and data insight consuption for organizations needs.

DATAVIRTUALIZATIONDATACOMPLIANCE PLATFORM ARCHITECTURE

The data compliance data virtualization platform architecture provides the landscape and layout of standard components required for more robust and secure compliant platform.

There are various building blocks of compliance platform are as follows:

- Region/ conutry specific data centers (Master & Worket Nodes)
- Enterprise-wide data virtualization platform
- Compliance reporting service
- User authorization
- Data & User policy engine
- Infrastructure gateways & firewalls
- User authentication

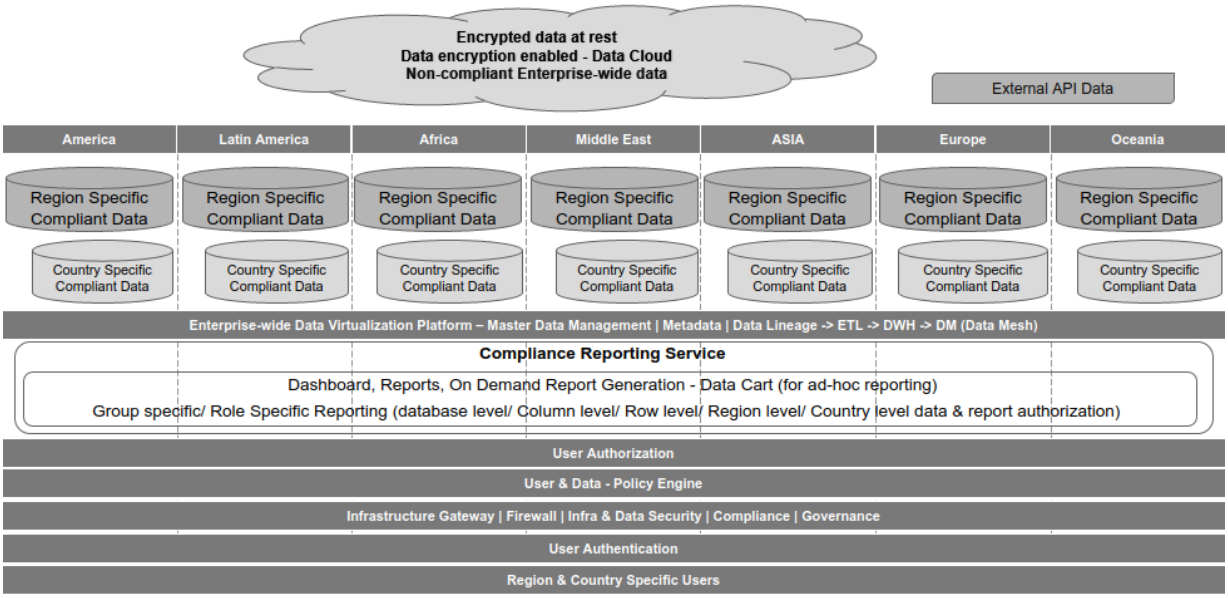

Figure 2: Data Compliance – Data Virtualization Platform Architecture

Tools and technologies involved in Data Virtualization Platform

On-premise databases: Oracle, SQL server, Big Data, PostgreSQL, IBM DB2...

Un-structed/ No SQL databases: MongoDB, Cassendra, DynamoDB, Hbase...

On-premise ETL tools: Informatica, IBM Datastage, talend, Oracle integrator, Pentaho...

Cloud databases: Snowflake, Azure bds, GCP dbs, AWS dbs, DataStack...

Gateways & Firewalls technologies

CICD pipelines: Jenkin, Git, Maveen...

BI tools: Power BI, Tableau, Qlik, Tibco Spotfire...

Data virtualization platforms

Actifio	Actifio Virtual Data Pipeline
AtScale	Intelligent Data Virtualization
Adata	CData Driver Technologies
Datameer	Datameer Spotlight
Data Virtuality	Data Virtuality Platform
Denodo	Denodo Platform

IBM	IBM Cloud Pak for Data
Informatica	Informatica PowerCenter
Oracle	Oracle Data Service Integrator
RedHat	Red Hat JBoss Data Virtualization
SAP	SAP HANA
SAS	SAS Federation Server
Stone Bond	Stone Bond Enterprise Enabler
Tibco	TIBCO Data Virtualization
Delphix	Delphix DevOps Data Platform
Vmware	VMware vCloud Director
AWS	AWS Glue

CONCLUSION

Data compliance - data virtualization framework can provide significant benefits to organizations, including manage their compliance obligations, manage risk, improve efficiency, gain competitive advantage, and continuously improve their compliance practice process, reduced risk of legal and regulatory penalties, increased customer trust and loyalty, improved data security, enhanced reputation and brand value, and increased operational efficiency. By implementing a data compliance framework using data virtualization technologies, organizations can ensure that they are operating within legal and regulatory requirements confined by regulatory authority, and are well positioned to achieve their organizations objective by utilizing local and global data sets. Data virtualization can play an vital role in data compliance by providing a flexible, robust and scalable approach to managing and protecting sensitive data. By leveraging data virtualization technology, organizations can enhance data security, data masking, data lineage, data quality, and compliance reporting.

REFERENCES

1. Adams, O. (2022, Dec 8). Privacy Laws in Different Countries and How to Comply With Them. Websitepolicies. https://www.websitepolicies.com/blog/privacy-laws-in-different-countries
2. Dilmegani, C. (2020, Jan 26). What is Data Virtualization? Benefits, Case Studies & Top Tools. research.aimultiple. https://research.aimultiple.com/data-virtualization/
3. Josyula, V. Orr, M, Page, G. (2012). Cloud Computing Automating the Virtualized Data Center. Cisco Press publication. p. 35-113
4. Kashdaran, A. (2020, Jul 30). What Is Data Compliance (Regulations And Standards). Incorporated. https://incorporated.zone/data-compliance/
5. Kosutic, D. (n. d.). Laws and regulations on information security and organizations continuity by country. Advisera. https://advisera.com/27001academy/knowledgebase/laws-regulations-information-security-organizations-continuity/
6. Marker, A. (2019, Aug 6). Regulatory Compliance 101 for organizations Managers. Smartsheet. https://www.smartsheet.com/content/regulatory-compliance-for-organizations-managers
7. Pohlman, M. (2008). Oracle Identity Management Governance, Risk, and Compliance Architecture. 3rd Edition. Chap 3. p. 41-46
8. powerdms. (2020, Dec 22). What Is Regulatory Compliance and Why Is It Important. Powerdms. https://www.powerdms.com/policy-learning-center/what-is-regulatory-compliance-and-why-is-it-important
9. proofpoint. (n. d.). What is IT Compliance. https://www.proofpoint.com/us/threat-reference/it-compliance
10. Quintero, D. Ahmad, F. Dominguez, S. Pontes, D. Rodriguz, C. (2019). Managing Security and Compliance in Cloud or Virtualized Data Centers Using IBM PowerSC. Redbook publication. p. 91-112
11. Rick, F. (2012). Data Virtualization for organizations Intelligence Systems Revolutionizing Data Integration for Data Warehouses. Elsevier. p. 1-266
12. Rick, F. (2018). Denodo Architecting the Multi-Purpose Data Lake with Data Virtualization. A Whitepaper. Denodo publication. P. 4-17
13. Unctad. (2021, Dec 14). Data Protection and Privacy Legislation Worldwide. unctad.org. https://unctad.org/page/data-protection-and-privacy-legislation-worldwide
14. Zhao, J. (2023, Feb 13). Data Compliance: What You Need to Know. Hyperproof. https://hyperproof.io/resource/data-compliance/

Abbreviations

DCIA	Digital Charter Implementation Act	
PDBD	Personal Data Protection Bill	India
PPA	Privacy Protection Authority	Israel
PPI	Protection of Personal Information	Japan
LGPD	Lei Geral de Proteçao de Dados	Brazil
PIPL	Personal Information Protection Law	China
EDPB	European Data Protection Board	EU
GDPR	General Data Protection Regulation	USA
CCPA	California Consumer Privacy Act	USA
POPIA	Protection of Personal Information Act	South Africa
PIPEDA	Personal Information Protection and Electronic Documents Act	Canada
NIST	National Institute of Standards and Technology	
HIPAA	Health Insurance Portability and Accountability Act	
CCPA	California Consumer Privacy Act	USA
PCI DSS	Payment Card Industry Data Security Standard	USA
SOX	Sarbanes-Oxley Act	USA
FedRAMP	Federal Risk and Authorization Management Program	
SHIELD	Stop Hacks and Improve Electronic Data Security	
CMMC	Cybersecurity Maturity Model Certification	