Data Architecture: Guiding Principles

A foundational value that guides and impacts decision-making and behaviour, providing a strategic roadmap for optimizing enterprise data management, governance, and utilization.

Classification:

EA Principle: Data Platform Centrality and SSOT (Single Source of Truth)

Enterprise Architecture and system designing must ensure that the data platform serves as the central hub for all data-related activities within the enterprise. All systems and processes should be designed to integrate with and leverage the capabilities of the data platform for data storage, processing, analysis, and consumption in as much **real time** as possible.

For data producers, it serves as the foundation for maintaining data integrity and coherence throughout the organization. By aligning with SSOT principles, data producers ensure that the data they generate or contribute to the central data platform remains consistent, accurate, and reliable.

All data used for decision-making, reporting, and analysis within the organization should originate from or be synchronized with the Data Platform to ensure consistency, accuracy, and reliability across all data-consuming systems.

The Single Source of Truth principle is foundational to data governance and data management practices within enterprises, providing a framework for establishing data quality standards, access controls, and data lineage tracking.

Guiding Principle #1, Data is a shared Asset.

**Collaborative Data Ownership:**

* There should be a shared responsibility for data across different departments and teams.
* Roles and accountabilities for data stewardship and management must be documented.

**Cross-Functional Collaboration:**

* Collaboration between various business units, IT, and analysts.
* Data solutions must not be built in isolation, it should cover a broader use case.
* Data Literacy.

**Data Sharing Policies:**

* Policies and standards for sharing data across teams and departments.
* Data sharing practices align with security, privacy, and compliance requirements.

**Centralized Data Repository:**

* Centralized data repository or data lake to facilitate shared access.
* Mechanisms for controlled access to prevent unauthorized usage.

**Data Transparency:**

* Transparency regarding the availability and quality of data.
* Implement Data Observability.
* Documentation and metadata to aid understanding and usage.

**Data Catalogs:**

* Centralized data catalog that indexes and describes available datasets.
* Enable users to discover and access relevant data easily.
* Standardized Formats and Definitions:
* Adopt standardized data formats and definitions to enhance interoperability.
* Ensure consistency in naming conventions and terminology across the organization.

**Data Sharing Agreements:**

* Establish formal agreements for data sharing, outlining terms, responsibilities, and expectations.
* Include considerations for data quality, security, and compliance in these agreements.

Guiding Principle #2, Usability and Accessibility.

* Prioritize making data accessible to a wide range of users, including business analysts, decision-makers, data scientists, and other stakeholders.
* Enable fast paced development and rapid deployment for changing user needs.
* Democratize access to data and analytics capabilities, encourage experimentation, exploration, and innovation with data.
* Support self-service analytics capabilities that enable users to access and analyse data independently, without relying on IT or data specialists.
* Ensuring that data is understandable and interpretable to users. This involves providing clear metadata, data dictionaries, and documentation that describe the meaning, context, and lineage of data attributes.
* Data Literacy: By providing user-friendly interfaces and educational resources, organizations can empower users to develop the skills and confidence needed to work effectively with data.
* Ensure that data access controls, privacy protections, and security measures are implemented in a way that does not hinder usability or accessibility for authorized users.

Guiding Principle #3, Data Security and Privacy

**Role-Based Access Control (RBAC):**

* Implement RBAC to assign permissions based on users' roles and responsibilities.
* Define roles such as admin, developer, QA user, with specific access levels. (include PII)
* Regularly review and update role assignments as personnel changes occur.

**Least Privilege Principle:**

* Apply the principle of least privilege to grant users the minimum access needed to perform their tasks.
* Avoid giving broad, unnecessary permissions to reduce the risk of unauthorized access.

**Access Policies and Procedures:**

* Establish clear access policies and procedures.
* Document the process for requesting, granting, modifying, and revoking access.
* Include approval workflows and proper documentation for auditing purposes.

**Authentication and Authorization:**

* Implement strong authentication mechanisms, such as AD based logging, Key Pair Authentication, multi-factor authentication (MFA).
* Ensure that users are authenticated before granting access and use authorization mechanisms to control what actions they can perform.

**Data Encryption:**

* Encrypt sensitive data at rest and in transit to protect it from unauthorized access.
* Use industry-standard encryption algorithms and regularly update cryptographic protocols.

**Audit Trails and Monitoring:**

* Implement comprehensive audit trails to log user activities.
* Regularly monitor and analyze logs for suspicious or anomalous behavior.
* Set up alerts for potential security incidents.

**Data Masking and Amendment:**

* Apply data masking and amendment techniques to hide sensitive information from unauthorized users.
* Protect confidential data by displaying only the necessary portions based on user permissions.

**Secure APIs and Data Transmission:**

* Ensure that APIs and data transmissions are secure by using HTTPS and other secure protocols.
* Implement API access controls and authentication mechanisms.

**Regular Security Audits and Assessments:**

* Conduct regular security audits and assessments to identify vulnerabilities.
* Perform penetration testing to simulate real-world attacks and strengthen defences.
* Address identified weaknesses promptly.

**Training and Awareness:**

* Provide ongoing security training for employees to raise awareness about best practices and potential threats.
* Emphasize the importance of safeguarding access credentials and sensitive information.

**Data Classification:**

* Classify data based on sensitivity and define appropriate access controls.
* Clearly communicate data classification policies to users.

**Incident Management:**

* Develop and maintain an incident response plan to address security breaches.
* Define roles and responsibilities for handling security incidents promptly and effectively.

**Regular Access Reviews:**

* Conduct regular access reviews to ensure that user permissions align with their roles and responsibilities.
* Remove or modify access for users who no longer require it.

Guiding Principle #4, Data Integration.

It involves combining data from disparate sources, formats, and systems to provide a unified view of the organization's data assets.

* Unified Data View: Data Integration allows organizations to create a unified view of data by combining information from multiple sources, such as databases, applications, files, and APIs.
* Frameworks – Data Ingestion framework, Transformation framework, well-articulated and definite business rules, Reporting and Data consumption Framework embracing Security and Privacy principles.
* Real time Insights: Real-time data integration is essential for applications and use cases that require immediate access to the latest information, such as customer support, fraud detection, and monitoring systems.
* Tactical changes and roadmap for strategic migration.
* Support for Analytics and Reporting: Integrated data serves as a foundation for analytics and reporting initiatives, providing a comprehensive dataset for analysis and visualization.
* Cross-Functional Collaboration: Data Integration promotes cross-functional collaboration by enabling different departments and teams to access and share data seamlessly.

Guiding Principle #5, Interoperability, Scalability and Flexibility.

**Interoperability:** This principle emphasizes the importance of ensuring that different systems, applications, and data sources can seamlessly exchange and use data.

**Scalability:** Scalability refers to the ability of the data architecture to accommodate growing data volumes, user demands, and processing requirements without significant performance degradation or system redesign.

**Flexibility:** Flexibility involves designing data architecture in a way that allows for easy adaptation to changing business needs, technological advancements, and data requirements.

++ Business Continuity

++ Disaster Management

Guiding Principle #6, Data Quality.

* **Accuracy:** Data should accurately reflect the real-world entities or phenomena it represents. Accuracy ensures that data is free from errors, discrepancies, or inconsistencies that could lead to incorrect conclusions or decisions. Data accuracy is essential for maintaining trust and confidence in data-driven insights and analyses.
* **Completeness:** Data completeness refers to the degree to which all relevant data is captured and available for analysis or decision-making purposes. Complete data sets contain all necessary information and attributes required to support specific business processes, analytics, or reporting requirements. Incomplete data can lead to gaps in understanding and hinder the ability to derive meaningful insights.
* **Consistency:** Consistency ensures that data is uniform and standardized across different sources, systems, or time periods. Consistent data adheres to predefined standards, formats, and definitions, facilitating seamless integration, analysis, and comparison of data across the organization. Inconsistent data can result in confusion, errors, and inaccuracies in reporting and analysis.
* **Relevance:** Data relevance pertains to the alignment of data with the specific needs, objectives, and requirements of the business or user. Relevant data is timely, appropriate, and applicable to the context in which it is used, enabling stakeholders to make informed decisions and derive actionable insights. Irrelevant or outdated data adds noise and complexity to data analysis, diminishing its value and effectiveness.
* **Timeliness:** Timeliness refers to the currency and freshness of data, ensuring that it is up-to-date and reflects the most recent information available. Timely data enables organizations to respond quickly to changing conditions, identify emerging trends, and seize opportunities in a rapidly evolving environment. Outdated or stale data can lead to missed opportunities and inaccurate decision-making.
* **Trustworthiness:** Trustworthiness encompasses the reliability, integrity, and credibility of data. Trustworthy data is trustworthy, transparent, and verifiable, instilling confidence in its accuracy, completeness, and validity. Establishing trust in data is essential for building confidence among stakeholders and ensuring that data-driven decisions are reliable and credible.

Guiding Principle #7, Data Lifecycle Management.

Implement efficient Data Lifecycle Management –

* Metadata Management
* Data Dictionary
* Data Lineage
* Change Capture
* Data Lineage
* Data Catalog
* Audit Trails
* Reconciliation
* DQ checks
* Observability

Guiding Principle #8, Operational Excellence.

* **Architecture Change Management.**

Process that governs the evolution of an organization's architecture over time. It involves managing proposed changes to the architecture, assessing their potential impact, and ensuring that they align with the organization's strategic objectives and architectural principles.

* **Strategic Alignment:**

Operational Excellence initiatives are aligned with the organization's strategic objectives and goals.

* **Standardization and Best Practices:**

standardization and the adoption of best practices across the organization.

* **Data-Driven Decision Making:**

data and metrics to drive decision-making and measure performance.

* **Lean Principles:**

focus on eliminating rework, optimizing workflows, and maximizing value, resulting in streamlined processes and reduced costs.

* **Risk Management and Resilience**

Proactive risk management practices to identify, assess, and mitigate operational risks.

* **Innovation and Agility:**

Innovation and agility to adapt to changing market dynamics, customer needs, and technological advancements.