Enterprise Data Strategy & Governance Framework

Executive Summary

This framework establishes comprehensive data governance and technical architecture strategy to ensure data integrity, security, and strategic value across enterprise systems. The strategy addresses organizational governance, technical implementation, and operational processes to support business decision-making and regulatory compliance.

Section 1: Data Governance Framework

1.1 Data Governance Structure

Data Governance Council

- **Chief Data Officer (CDO)**: Executive sponsor and strategic decision authority
- **Data Governance Committee**: Cross-functional leadership team
- **Data Stewards**: Operational data quality and compliance oversight
- **Business Data Owners**: Domain expertise and business rule definition

Data Ownership Model

Data Domain	Data Owner	Data Steward	Business Sponsor
Customer Data	VP Marketing	Marketing Operations Manager	Chief Marketing Officer
Financial Data	CFO	Financial Systems Manager	Chief Financial Officer
Product Data	VP Product	Product Data Manager	Chief Product Officer
Operational Data	VP Operations	Business Systems Analyst	Chief Operations Officer
Employee Data	CHRO	HR Systems Manager	Chief Human Resources Officer

1.2 Data Quality Standards

Quality Metrics Framework

Completeness Standards:

- Customer contact data: 98% (critical for marketing and support)
- Product specifications: 95% (required for e-commerce operations)
- Financial transaction data: 100% (regulatory compliance requirement)
- Employee records: 97% (HR and payroll operations)

Accuracy Requirements:

- Customer demographics: 96% accuracy validated quarterly
- Product pricing: 99.9% accuracy (revenue impact)
- Inventory levels: 98% accuracy (operational efficiency)
- Financial reporting data: 99.8% accuracy (regulatory compliance)

Consistency Standards:

- Standardized customer naming conventions across all systems
- Unified product categorization taxonomy
- Consistent date/time formats (ISO 8601 standard)
- Standardized currency and measurement units

Timeliness Requirements:

- Customer updates: Real-time synchronization within 15 minutes
- Product information: Updated within 2 hours of source changes
- Financial data: End-of-day batch processing with next-day availability
- Inventory data: Real-time updates for critical stock levels

1.3 Data Classification and Security

Classification Framework

Public Data: Marketing materials, published financial reports, public website content

- Access: Unrestricted
- Retention: 7 years
- Protection: Standard backup procedures

Internal Data: Employee directories, internal metrics, operational reports

- Access: Employee access with role-based restrictions
- Retention: 5 years active, 2 years archived
- Protection: Standard encryption and access controls

Confidential Data: Customer PII, detailed financial records, strategic plans

- Access: Need-to-know basis with manager approval
- Retention: Regulatory requirements (varies by data type)
- Protection: Encryption at rest and in transit, audit logging

Restricted Data: Legal documents, M&A information, employee personal data

- Access: C-level approval required
- Retention: Legal hold requirements
- Protection: Multi-factor authentication, data loss prevention

1.4 Data Lifecycle Management

Creation and Ingestion

- **Data Entry Standards**: Mandatory field validation, format checking
- **Source System Integration**: Automated data capture with quality checks
- **Third-Party Data**: Vendor data quality agreements and validation processes
- **User Training**: Data entry best practices and quality awareness

Storage and Maintenance

- **Primary Storage**: Production systems with real-time access
- **Archive Storage**: Long-term retention for compliance and historical analysis
- **Backup Procedures**: Daily incremental, weekly full backups with offsite storage
- **Data Refresh**: Regular updates from authoritative sources

Usage and Access Control

- **Role-Based Access**: Granular permissions based on job function
- **Usage Monitoring**: Audit trails for data access and modifications
- **API Governance**: Rate limiting and usage tracking for system integrations
- **Analytics Access**: Self-service analytics with governed data sets

Disposal and Retention

- **Retention Schedules**: Legal and business requirements by data category
- **Secure Deletion**: Cryptographic erasure and certificate of destruction
- **Legal Hold**: Litigation and regulatory investigation procedures
- **Privacy Compliance**: Right to erasure and data portability requirements

Section 2: Data Architecture Strategy

2.1 Current State Assessment

Existing Data Landscape

Core Business Systems:

- **CRM Platform**: Salesforce Sales Cloud (customer data, opportunity management)
- **ERP System**: SAP S/4HANA (financial, procurement, inventory management)
- **E-commerce Platform**: Shopify Plus (online orders, product catalog, customer accounts)
- **Marketing Automation**: HubSpot (lead management, campaign tracking)
- **Customer Support**: Zendesk (ticket management, customer interactions)

Data Storage and Analytics:

- **Data Warehouse**: Snowflake (reporting, business intelligence)
- **Business Intelligence**: Tableau (dashboards, self-service analytics)
- **File Storage**: Box (document management, collaboration)

Integration Challenges:

- 15+ point-to-point integrations with manual data transfers
- Inconsistent data formats across systems
- No real-time synchronization capabilities
- Limited data governance and quality monitoring
- Siloed analytics with conflicting business metrics

2.2 Target State Architecture

Master Data Management Strategy

Customer Master: Salesforce as single source of truth for customer information

- Consolidated customer profiles from all touchpoints
- Real-time synchronization to downstream systems
- Data quality monitoring and automated cleansing

Product Master: SAP as authoritative source for product information

- Centralized product catalog with specifications and pricing
- Automated distribution to e-commerce and marketing systems
- Version control and change management processes

Financial Master: SAP as system of record for financial transactions

- General ledger integration with all revenue-generating systems
- Real-time financial reporting and reconciliation
- Compliance and audit trail maintenance

Data Platform Architecture

Cloud-First Strategy: AWS as primary cloud platform

- **Data Lake**: Amazon S3 for raw data storage and exploration
- **Data Warehouse**: Snowflake for structured analytics and reporting
- **Real-Time Processing**: Amazon Kinesis for streaming data
- **Batch Processing**: AWS Glue for ETL operations

Integration Hub: MuleSoft Anypoint Platform

- **API Management**: Centralized API governance and monitoring
- **Data Transformation**: Standardized mapping and cleansing rules
- **Error Handling**: Automated retry and exception management
- **Security**: OAuth 2.0 and API key management

2.3 Integration Patterns and Technology Stack

Real-Time Integration Patterns

Event-Driven Architecture: Apache Kafka for high-volume data streaming

- Customer behavior events from e-commerce platform
- Inventory updates from warehouse management systems
- Financial transaction processing and reconciliation

API-First Integration: RESTful APIs for system-to-system communication

- Customer data synchronization between CRM and support systems
- Product information distribution to marketing and e-commerce platforms
- Order processing workflow automation

Batch Processing Patterns

ETL Workflows: Scheduled data processing for non-critical updates

- Historical data migration and archive processing
- Bulk data quality assessment and cleansing
- Regulatory reporting and compliance data preparation

Data Replication: Database synchronization for reporting and analytics

- Daily snapshots of transactional systems
- Data warehouse refreshes for business intelligence
- Backup and disaster recovery data maintenance

Technology Stack Components

Data Processing: Apache Spark for large-scale data transformation **Monitoring and Observability**: DataDog for infrastructure and data pipeline monitoring **Data Catalog**: Apache Atlas for metadata management and data discovery **Security**: HashiCorp Vault for secrets management and encryption key rotation

2.4 Implementation Roadmap

Phase 1: Customer Data Consolidation (Months 1-6)

Objectives: Establish single customer master and eliminate data silos

Deliverables:

- Salesforce configured as customer master data hub
- Real-time API connections from HubSpot, Zendesk, and Shopify
- Customer data quality monitoring dashboard
- Automated data cleansing and standardization processes

Success Criteria:

- 98% customer data completeness across all systems
- Customer data synchronization within 15 minutes
- 50% reduction in customer data discrepancies

Resources Required:

- 2 Salesforce developers (6 months)
- 1 Integration architect (6 months)
- 1 Data analyst (3 months)
- Budget: $275,000

Key Risks and Mitigation:

- Legacy system API limitations → Develop custom middleware solutions
- Data quality issues in source systems → Implement data cleansing workflows
- User adoption resistance → Comprehensive training and change management

Phase 2: Real-Time Order Processing (Months 7-12)

Objectives: Enable real-time order-to-cash process automation

Dependencies:

- Phase 1 customer master completion
- MuleSoft platform implementation
- Kafka infrastructure deployment

Deliverables:

- Real-time order processing from e-commerce to ERP
- Automated inventory updates and availability checking
- Financial transaction posting and reconciliation
- Customer notification and tracking systems

Success Criteria:

- Order processing time reduced from 4 hours to 15 minutes
- 99.5% order accuracy rate
- Real-time inventory visibility across all channels

Resources Required:

- 3 Integration developers (6 months)
- 1 Solution architect (6 months)
- 1 DevOps engineer (4 months)

- Budget: $420,000

Go/No-Go Decision Criteria:

- Customer data accuracy >97% sustained for 30 days
- API performance <2 seconds response time
- Zero critical system outages during Phase 1

Phase 3: Advanced Analytics Platform (Months 13-18)
Objectives: Implement self-service analytics and predictive capabilities

Dependencies:

- Phases 1 and 2 successful completion
- Data warehouse optimization
- Analytics user training completion

Deliverables:

- Real-time analytics dashboard for executives
- Self-service business intelligence for operations teams
- Predictive analytics for inventory and demand planning
- Automated reporting and alerting systems

Success Criteria:

- 80% of reports generated through self-service tools
- 25% improvement in demand forecasting accuracy
- 90% user satisfaction with analytics capabilities

Resources Required:

- 2 Data engineers (6 months)
- 1 Analytics architect (6 months)
- 1 Business intelligence developer (6 months)
- Budget: $380,000

Risk Mitigation Strategies:

- Pilot program with limited user group before full rollout
- Comprehensive user training and documentation
- Fallback procedures for critical reporting processes

2.5 Success Metrics and Monitoring

Key Performance Indicators

Data Quality Metrics:

- Overall data completeness: Target 96% across all systems
- Data accuracy rate: Target 98% for critical business data
- Data synchronization time: Target <30 minutes for all integrations

Operational Efficiency:

- Reduction in manual data entry: Target 75% decrease
- Time to generate reports: Target 60% improvement
- System integration errors: Target 90% reduction

Business Value Metrics:

- Customer 360 view completeness: Target 95%
- Decision-making speed improvement: Target 40%
- Regulatory compliance score: Target 100%

Monitoring and Governance

Technical Monitoring: Real-time monitoring of data pipeline health and performance **Business Monitoring**: Regular assessment of data quality and business value metrics **Governance Reviews**: Quarterly review of data governance policies and compliance **Continuous Improvement**: Monthly stakeholder feedback and process optimization

Conclusion

This Enterprise Data Strategy & Governance Framework provides the foundation for transforming organizational data capabilities from fragmented, siloed systems to a unified, governed, and strategically valuable data ecosystem. Success depends on sustained executive commitment, cross-functional collaboration, and disciplined execution of the phased implementation approach.