



CHAPTER 7 – DATA MANAGEMENT IN CLOUD COMPUTING

A decorative geometric pattern on the left side of the slide, featuring a grid of squares with various patterns: a blue square with a white circle, a blue square with concentric circles, a purple square with a white circle, a blue square with concentric circles, a purple square with a white circle, and a blue square with concentric circles.

AGENDA

1. Data Security in Cloud Computing
2. Data Location, Control, Transport etc.
3. Data Processing
4. Databases and Data stores.



OVERVIEW OF DATA MANAGEMENT IN CC

- Introduction to Data Management
- Importance of Data Management in Cloud



DATA MANAGEMENT

- **Data management** is the practice of collecting, storing, organizing, protecting, and analyzing data to derive valuable insights.
- **Key components:** Data governance, data quality, data security, data integration, data warehousing, and data mining.
- **The Evolution of Data Management:**
 - Traditional data management: On-premises data centers, challenges with scalability, cost, and disaster recovery.
 - Cloud data management: Benefits of cloud-based data management, such as scalability, cost-efficiency, and accessibility.



IMPORTANCE OF DATA MANAGEMENT

- **Data as a strategic asset:**
 - **Role in Decision-Making:** Data-driven decisions lead to better outcomes
 - **Innovation:** Leveraging data for innovation and competitive advantage
- **Data security and privacy:**
 - **Importance:** Protecting sensitive and personal information
 - **Risks:** Potential threats such as breaches, leaks, and unauthorized access
 - **Measures:** Encryption, access controls, and regular audits
- **Cost optimization:** Explain how effective data management can reduce storage and processing costs.



COMPLIANCE AND REGULATIONS

Industry Standards: Examples like GDPR, HIPAA, and CCPA

Importance: Ensuring legal compliance and avoiding penalties

Best Practices: Regular compliance audits and documentation

Note:

GDPR: General Data Protection Regulation

HIPAA: Health Insurance Portability and Accountability Act

CCPA: California Consumer Privacy Act



COST OPTIMIZATION

Cost Optimization

- **Benefits:** Reducing storage and processing costs
- **Techniques:** Efficient data storage, deduplication, and tiered storage solutions
- **Examples:** Cloud providers' cost management tools



CHALLENGES IN CLOUD DATA MANAGEMENT

- Data Volume and Velocity
- Managing large datasets
- Rapid data growth
- Data Quality
- Accuracy
- Consistency
- Completeness
- Data Security and Privacy



CHALLENGES *CONTINUED*

- Risks of data breaches
- Unauthorized access
- Data Governance
- Managing data across platforms
- Departmental data management complexities



THE IMPORTANCE OF DATA SECURITY

- **Data Breaches and Their Impact**
 - Financial losses
 - Reputational damage
 - Legal liabilities
- **The Shared Responsibility Model**
 - Roles of cloud providers
 - Roles of customers
 - Compliance Regulations



THE CIA TRIAD OF SECURITY

- **Confidentiality**
 - Protecting data from unauthorized access.
- **Integrity**
 - Ensuring data accuracy and completeness.
- **Availability**
 - Ensuring data is accessible when needed.



THREATS TO DATA SECURITY

- **Data breaches:** Unauthorized access to sensitive data.
- **Malware:** Malicious software designed to damage, disrupt, or gain unauthorized access to systems.
- **Ransomware:** A type of malware that encrypts data and demands a ransom for its release.
- **Phishing:** Fraudulent attempts to obtain sensitive information by disguising as a trustworthy entity.
- **DDoS attacks:** Distributed Denial of Service attacks that overwhelm a system with traffic to make it unavailable.



INSIDER THREATS:

- **Employees as a security risk:** Employees with access to sensitive data can intentionally or unintentionally compromise data security.
- **Cloud-Specific Threats:**
 - **Account hijacking:** Unauthorized access to cloud accounts.
 - **Data loss:** Permanent loss of data due to accidental deletion, corruption, or malicious actions.
 - **Unauthorized access:** Access by individuals or entities that do not have permission to use the data.



SECURITY CONTROLS

- **Preventive Controls:**
 - **Firewalls:** Act as barriers between trusted and untrusted networks, blocking unauthorized access.
 - **Access controls:** Restrict access to data and systems based on user roles and permissions.
 - **Intrusion prevention systems (IPS):** Monitor network traffic to detect and prevent malicious activities.
- **Detective Controls:**
 - **Intrusion detection systems (IDS):** Monitor network and system activities for suspicious behavior.
 - **Log analysis:** Review and analyze logs to identify and investigate security incidents.
 - **Anomaly detection:** Identify unusual patterns or behaviors that may indicate a security threat.
- **Corrective Controls:**
 - **Incident response plans:** Define procedures for responding to and managing security incidents.
 - **Data recovery:** Ensure the ability to restore data following a loss or corruption event.



CLOUD SECURITY BEST PRACTICES

- **Risk Assessment:** Identifying and prioritizing security risks.
- **Data Encryption:** Protecting data at rest and in transit.
- **Access Management:** Implementing strong authentication and authorization.
- **Regular Security Audits and Monitoring:** Identifying vulnerabilities and threats.
- **Employee Training and Awareness:** Building a security-conscious culture.



IDENTITY AND ACCESS MANAGEMENT (IAM)

- **Importance of IAM:** Controlling who can access cloud resources.
- **Key Components:**
 - Authentication
 - Authorization
 - Single sign-on (SSO)
- **Role-based Access Control (RBAC):** Assigning permissions based on roles.
- **Least Privilege Principle:** Granting only necessary permissions.



DATA ENCRYPTION

Encryption Methods:

- Symmetric encryption
- Asymmetric encryption

Key Management:

- Key generation
- Key distribution
- Key storage

Encryption at Rest and in Transit:

- Protecting data in storage
- Protecting data during transmission



SECURITY AS A SERVICE (SAAS)

- **Cloud Security Offerings:**
 - Intrusion Detection and Prevention Systems (IDPS)
 - Vulnerability Scanning
 - Data Loss Prevention (DLP)
- **Benefits of SaaS for Security:**
 - Scalability
 - Expertise
 - Cost-effectiveness



CLOUD SECURITY COMPLIANCE AND REGULATIONS

- **Industry-Specific Regulations:**
 - HIPAA
 - GDPR
 - PCI DSS
 - SOX
- **Compliance Challenges in the Cloud:**
 - Data Residency
 - Data Sovereignty
- **Best Practices for Compliance:**
 - Risk Assessment
 - Audits and Certifications



INCIDENT RESPONSE AND DISASTER RECOVERY

- **Incident Response Plan:** Steps to take in case of a security breach
- **Disaster Recovery Planning:** Protecting data and systems from failures
- **Business Continuity Planning:** Ensuring operations can continue after a disaster




EMERGING CLOUD SECURITY THREATS AND COUNTERMEASURES

- **Cloud-Native Threats:**
 - Serverless attacks
 - Container vulnerabilities
- **Emerging Security Technologies:**
 - Behavioral analytics
 - Machine learning for threat detection



DATA LAKES, BIG DATA

- **Data Lakes:** A vast repository of raw data in its native format.
 - **Benefits:** Cost-effective, flexible, and scalable.
 - **Challenges:** Governance, security, and processing.
- **Big Data and Hadoop**
 - **Big Data characteristics:** Volume, velocity, variety, veracity.
 - **Hadoop Distributed File System (HDFS):** Scalable storage for big data.
 - **MapReduce:** Distributed processing framework.



DATA WAREHOUSING & DATA LAKES: A COMPARISON

- **Key Differences:**
 - **Structure:** Data Warehouses (Structured data) vs. Data Lakes (Structured and Unstructured data)
 - **Purpose:** Data Warehouses (Reporting and analysis) vs. Data Lakes (Raw data storage for future processing)
 - **Use Cases:**
 - Data Warehouses: Business Intelligence (BI), OLAP
 - Data Lakes: Machine Learning (ML), Data Mining, Advanced Analytics



DATA GOVERNANCE AND QUALITY

- **The Importance of Data Governance**
- **Definition of data governance:** The overall management of the availability, usability, integrity, and security of the data used in an enterprise.
- **Benefits of effective data governance:** Improved decision-making, risk mitigation, regulatory compliance, and cost savings.
- **Data governance challenges in the cloud:** Data ownership, data security, data quality, and regulatory compliance.



DATA GOVERNANCE FRAMEWORK

- **Key components of a data governance framework:** Policies, standards, roles, and responsibilities.
- **Data stewardship:** Defining roles and responsibilities for data management.
- **Data ownership:** Identifying data owners and their accountability.
- **Data quality management:** Ensuring data accuracy, completeness, and consistency.



DATA ASSESSMENT

QUALITY

- **Key dimensions of data quality:** Accuracy, completeness, consistency, timeliness, validity, and uniqueness.
- **Data profiling techniques:** Identifying data characteristics and quality issues.
- **Data quality metrics:** Measuring data quality



DATA CLEANSING AND ENRICHMENT

- **Data cleansing process:** Correcting errors, handling missing values, and standardizing data formats.
- **Data enrichment techniques:** Adding context and value to data through external sources.
- **Data deduplication:** Identifying and removing duplicate records.



CLOUD DATA INTEGRATION AND INTEROPERABILITY

- **Data Integration Challenges in the Cloud**
 - **Heterogeneous data sources:** Handling different formats and structures.
 - **Data quality issues:** Ensuring data consistency and accuracy.
 - **Data security and privacy:** Protecting sensitive data during integration.



MISCELLANEOUS

- **Data Integration Patterns and Tools**
 - ETL vs. ELT: Comparison of data integration approaches.
 - Data integration tools: Cloud-based and on-premises options.
 - Data virtualization: Creating a unified view of data without physical integration.
- **Data Interoperability**
 - Standards and protocols: XML, JSON, REST, SOAP.
 - Data exchange formats: CSV, Excel, Parquet, Avro.
 - Master data management (MDM): Ensuring data consistency across systems.



DATA LOCATION

- **Importance of Data Location:**
 - Ensures optimal performance and user experience.
 - Affects data latency and access speed.
- **Factors Influencing Data Location:**
 - **Latency:** Proximity to users reduces response time.
 - **Regional Laws:** Legal requirements for data storage.
 - **Data Sovereignty:** Control over data by the host country.
- **Regulatory Requirements:**
 - GDPR: General Data Protection Regulation.
 - Other region-specific laws and regulations.




SECURING DATA FOR TRANSPORT

- **Encryption Methods**
 - Secure Protocols (HTTPS, TLS)
 - Best Practices for Data Transport



SCALABILITY IN CLOUD SERVICES

- Definition of Scalability
- Horizontal vs. Vertical Scaling
- Benefits of Scalability



LARGE SCALE DATA PROCESSING

- **Tools and Technologies (Hadoop, Spark)**