Module 5 - Data Storage and Operations

What is Data Storage and Operations

"**Data storage and Operations** includes the design, implementation, and support of stored data, to maximize its value throughout its lifecycle, from creation/acquisition to disposal"

Two sub-activities:

1. Database support
2. Database technology support

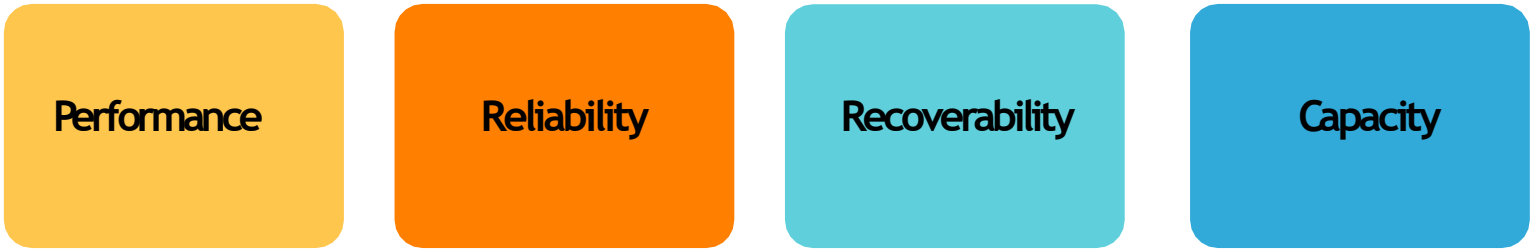
Source: DAMA-DMBOK2 Data Management

Benefits of good Data Storage Strategy

1. *Reduce capital expenses*
2. *Reduce operational expenses*
3. *Easier data management*
4. *Optimized resource utilization*
5. *Easier scalability*
6. *Better performance*
7. *Better user experience*

Data Storage activities

Data Storage Management key attributes



Data Storage Best Practices

Know your data needs

Tiered data storage

Disaster recovery plans

Consolidate Storage Pools

Utilize automated storage management tools

Module 6 - Data Security

What is Data Security

“**Data Security** is the process of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.”

Why is Data Security important

- The average total cost of a ransomware breach is $4.62 million, slightly higher than the average data breach of $4.24 million (IBM)
- The average per record (per capita) cost of a data breach increased by 10.3 percent from 2020 to 2021 (IBM)
- The average cost of a breach with a lifecycle over 200 days is $4.87 million (IBM)
- 39 percent of costs are incurred more than a year after a data breach (IBM)
- In 2021, the United States was the country with the highest average total cost of a data breach was at $9.05 million (IBM)
- 34 percent of data breaches in 2018 involved internal actors (Verizon)
- It took an average of 287 days to identify a data breach (IBM).

Source: https://www.varonis.com/blog/data-breach-statistics

Types of Data Security

Sources: https://www.varonis.com/blog/data-security

Data Security Risks

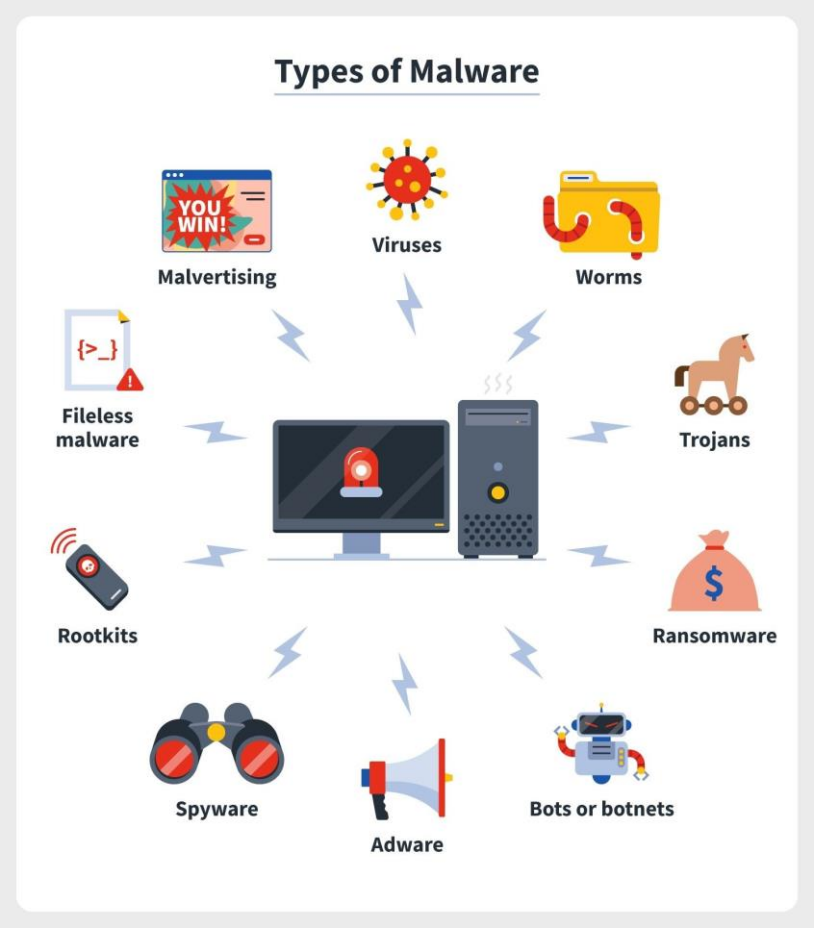
- Accidental Data Exposure
- Phishing
- Malware
- Insider Threats
- Password Attack
- Denial-of-Service (DOS)
- Man-in-the-Middle (MITM)
- SQL Injections
- Zero-day Exploit

Accidental Data Exposure

Phishing

Source: https://www.itgovernance.co.uk/phishing

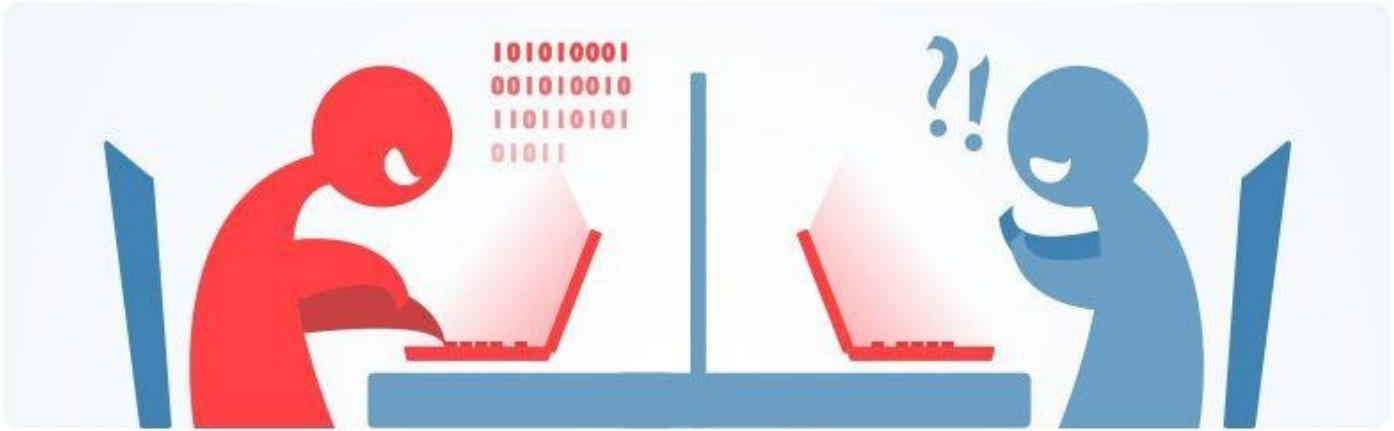
Malware



Source:
https://us.norton.com/internetsecurity-malware-types-of-malware.html

Insider Threats

Insider threat classification by CA Technologies



Malicious insiders

Intentionally use their access to sensitive data to harm the company

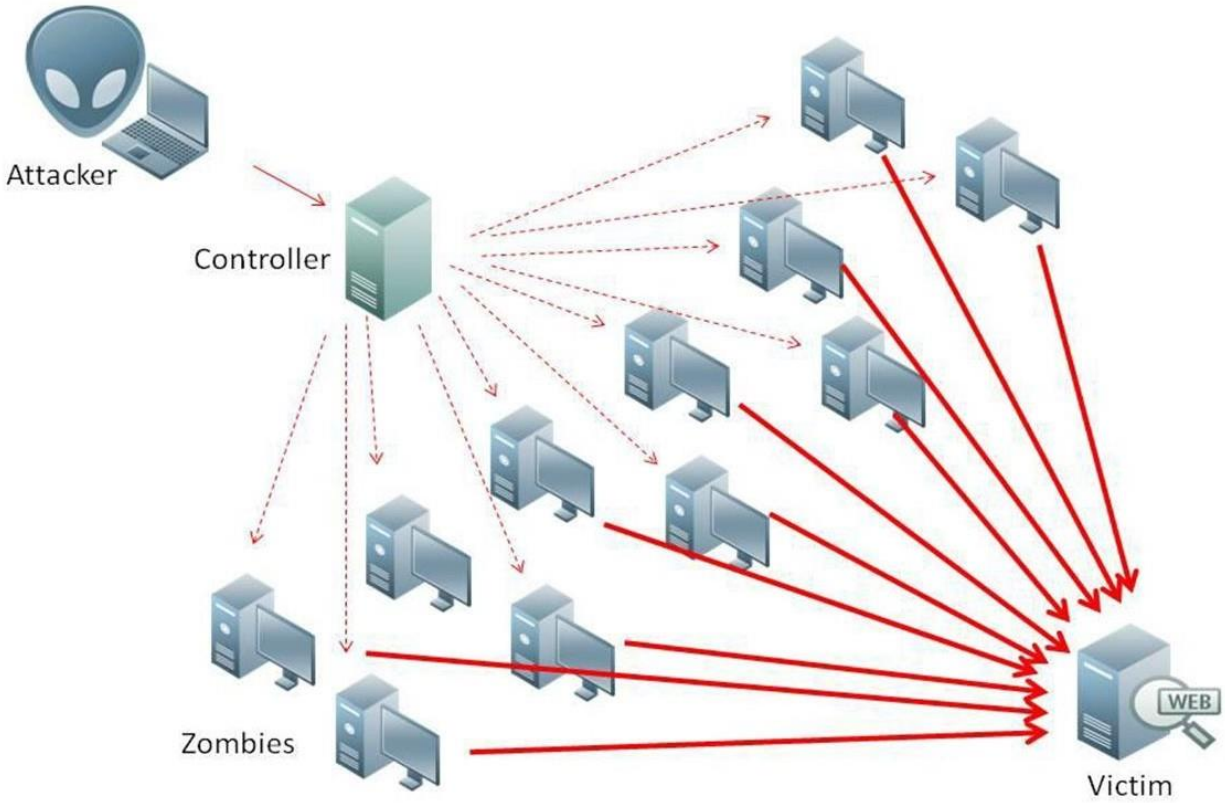
Inadvertent insiders

Cause damage to the company unintentionally

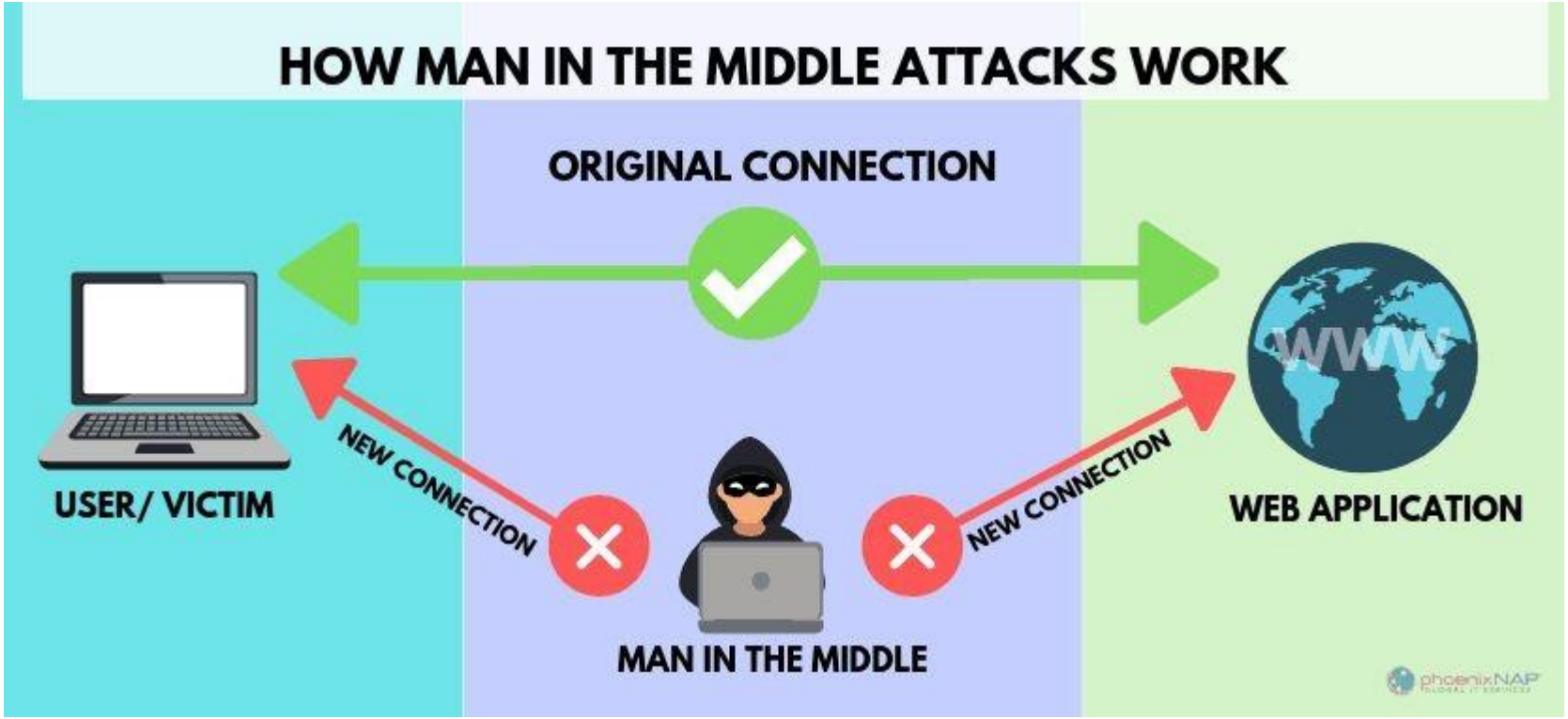
Source: https://www.ekransystem.com/en/blog/insider-threat-definition

Password Attacks

Denial-of-Service (DOS)



Man-In-The-Middle



SQL Injections

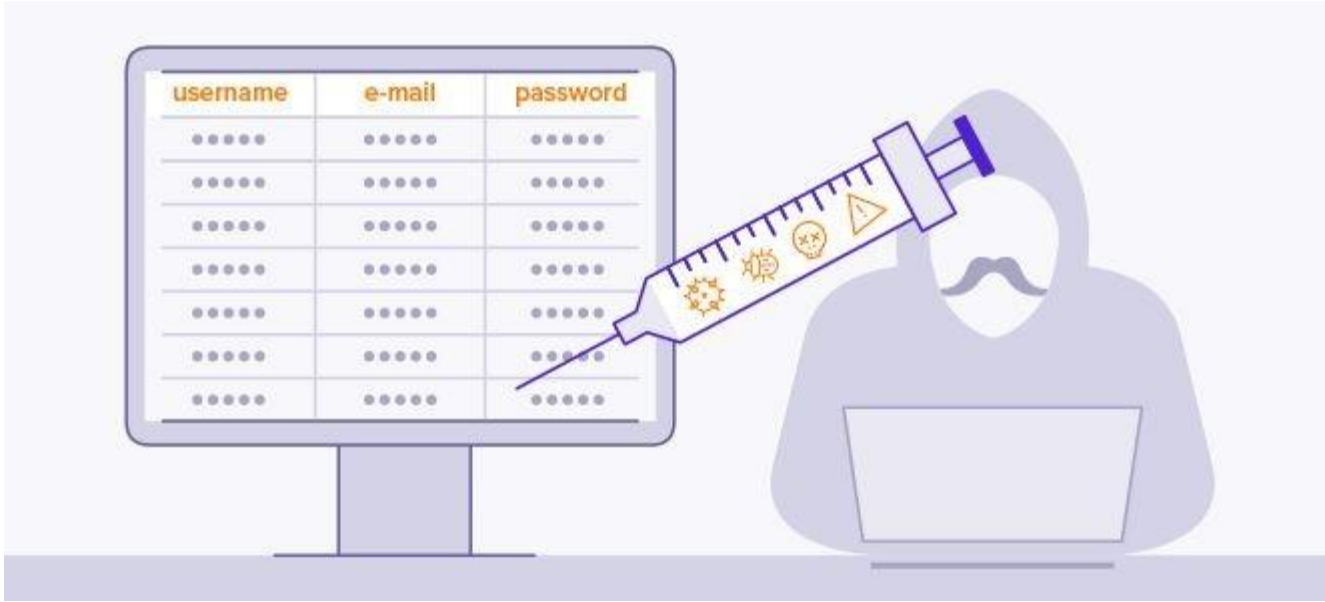


Image Source: https://www.avast.com/c-sql-injection

Zero-Day Exploit

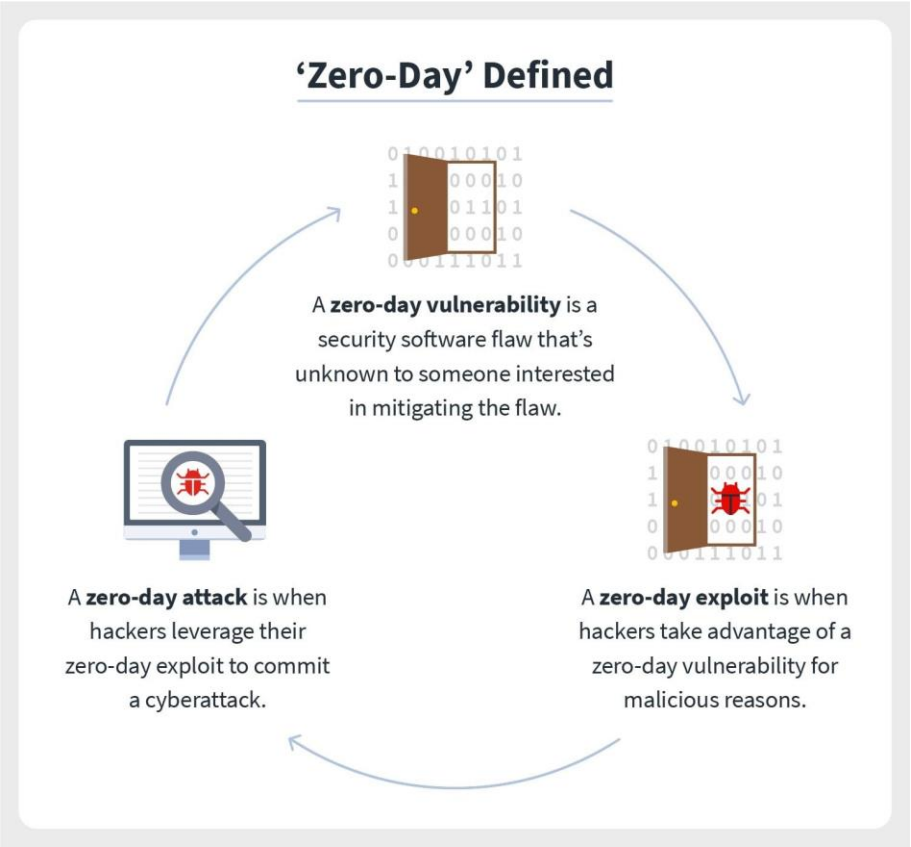


Image Source: https://us.norton.com/internetsecurity-emerging-threats-how-do-zero-day-vulnerabilities-work.html

9 Best Practices to secure your data

1. Employees education
2. Create Insider Threat Policies
3. Phishing Simulations
4. Backup data
5. Update Systems and Software
6. Utilize HTTPS
7. Maintain Compliance
8. Use multi-factor authentication
9. Employe latest secure coding practices

Module 7 - Data Integration

What is Data Integration

“**Data Integration** is the process of consolidating data from different sources into one, unified view for efficient data management”

Example of Data Integration

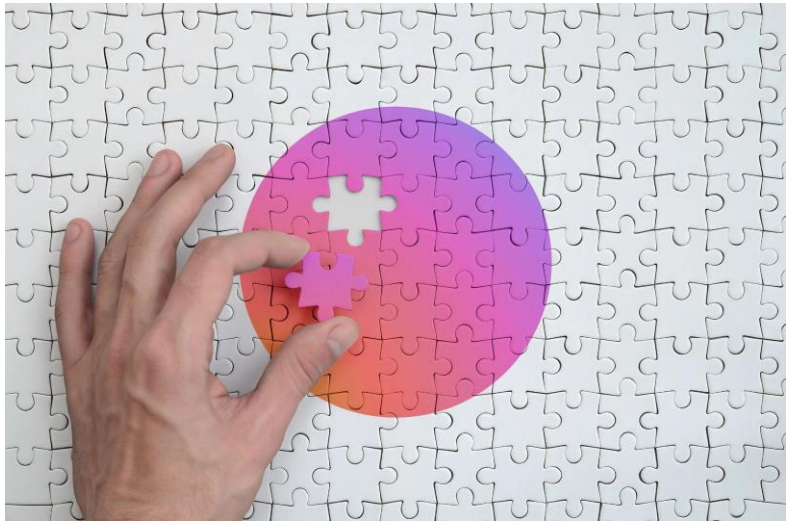
Company A uses the following:

- Salesforce for customer information and sales pipeline data
- External vendor for additional customer firmographics
- Internal database that tracks customer satisfaction ratings from surveys
- Internal financial system that tracks the sales revenue per customer
- Marketing department database on customer campaigns

Goal: Integration all the above sources into single source!

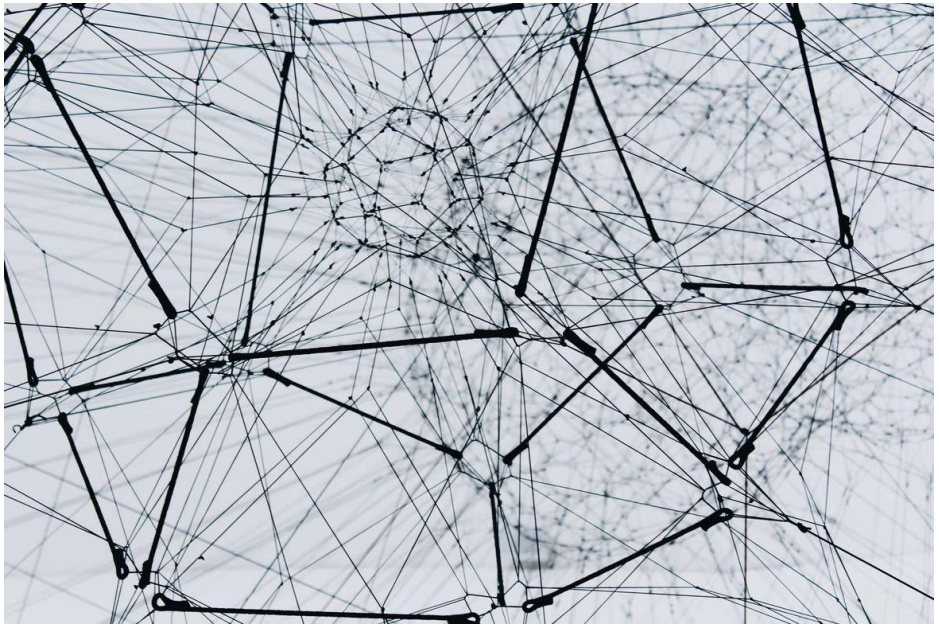
Importance of Data Integration

- Complete view of business intelligence, insights and analytics
- Increased efficiency and ROI - no need for manual data gathering
- Better employee, customer and partner experience
- Improves Collaboration
- Eliminates Data Silos
- Reduces errors
- Faster innovation, sales, time to market
- Improves Data Quality and Integrity



Techniques for Data Integration

- Manual Data Integration
- Middleware Data Integration
- Application Based Integration
- Uniform Access Integration
- Common Storage Integration (Data Warehousing)
- Data Virtualization



Manual Data Integration

Pros:

- Low cost
- Total control

Cons:

- Difficult to scale
- Human error

Source: https://www.talend.com/resources/data-integration-methods/

Middleware Data Integration

Pros:

- Better data streaming
- Easier access between systems

Cons:

- Maintenance
- Limited functionality

Source: https://www.talend.com/resources/data-integration-methods/

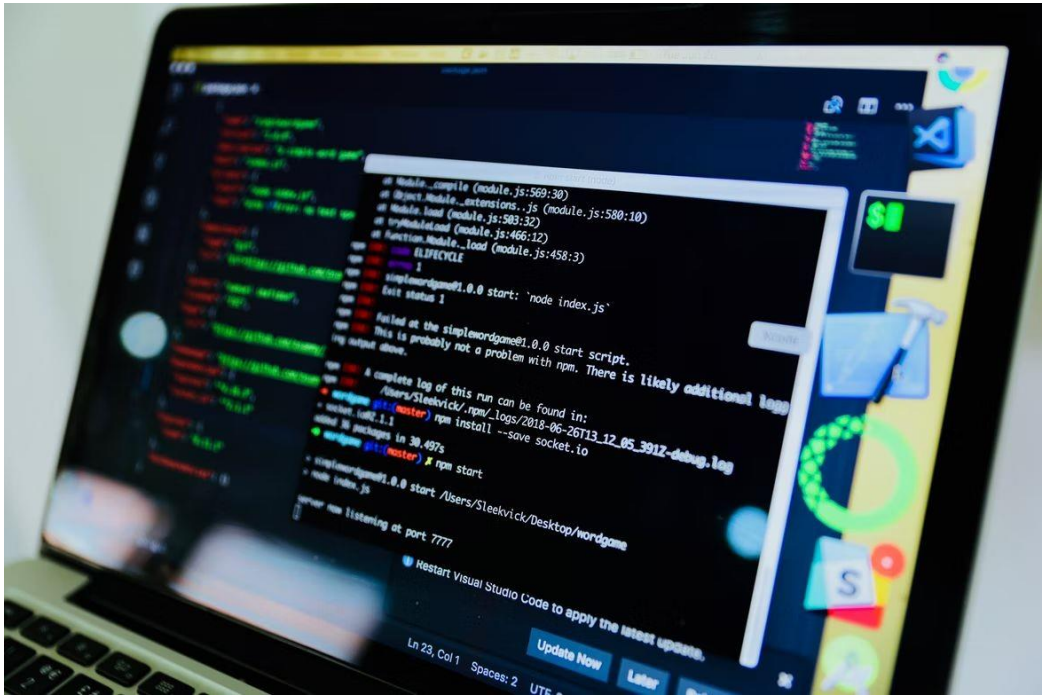
Application Based Integration

Pros:

- Simplified processes
- Easier information exchange
- Fewer resources used

Cons:

- Maintenance
- Inconsistent results
- Complicated setup
- Difficult data management



Source: https://www.talend.com/resources/data-integration-methods/

Uniform access Integration

Pros:

- Lower storage requirements
- Easier data access
- Simplified view of data

Cons:

- Data integrity challenges
- Strained systems

Source: https://www.talend.com/resources/data-integration-methods/

Common Storage Integration (Data Warehousing)

Pros:

- Reduced burden
- Increased data version management control
- Cleaner data
- Enhanced data analytics

Cons:

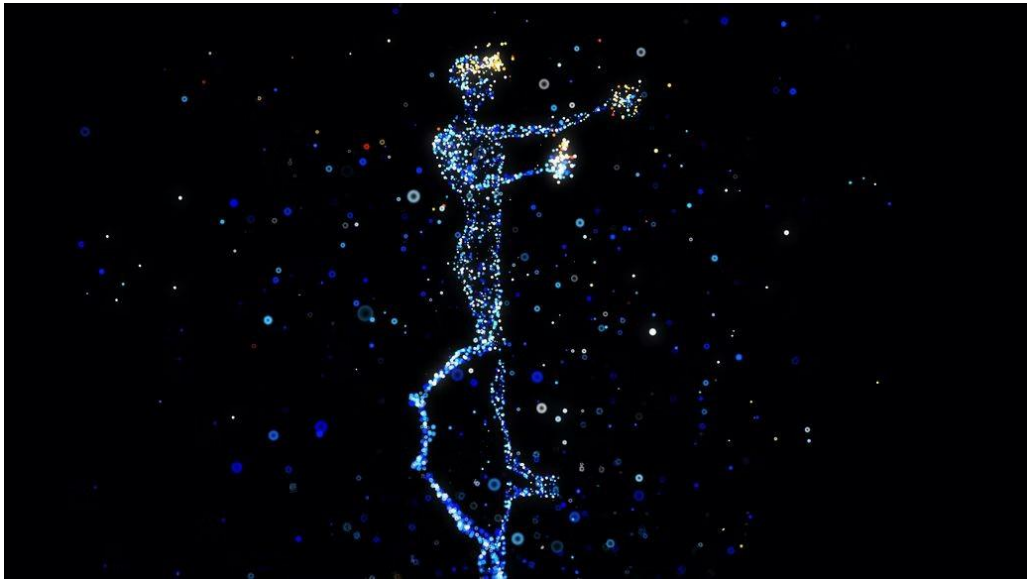
- Increased storage costs
- Higher maintenance costs

Source: https://www.talend.com/resources/data-integration-methods/

Data Virtualization

Pros:

- No need to move data
- Scalable
- No need to maintain data in multiple locations

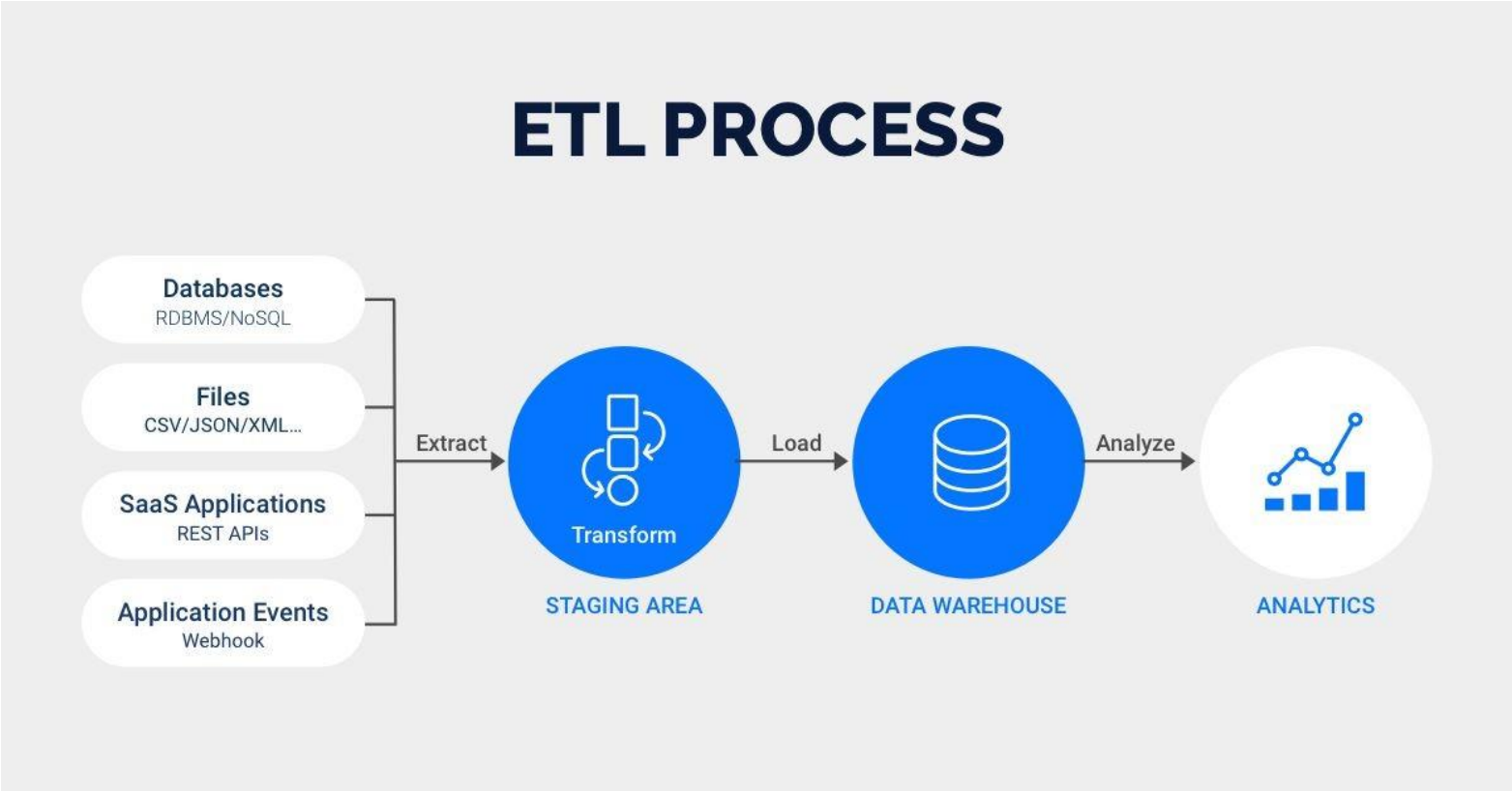


Cons:

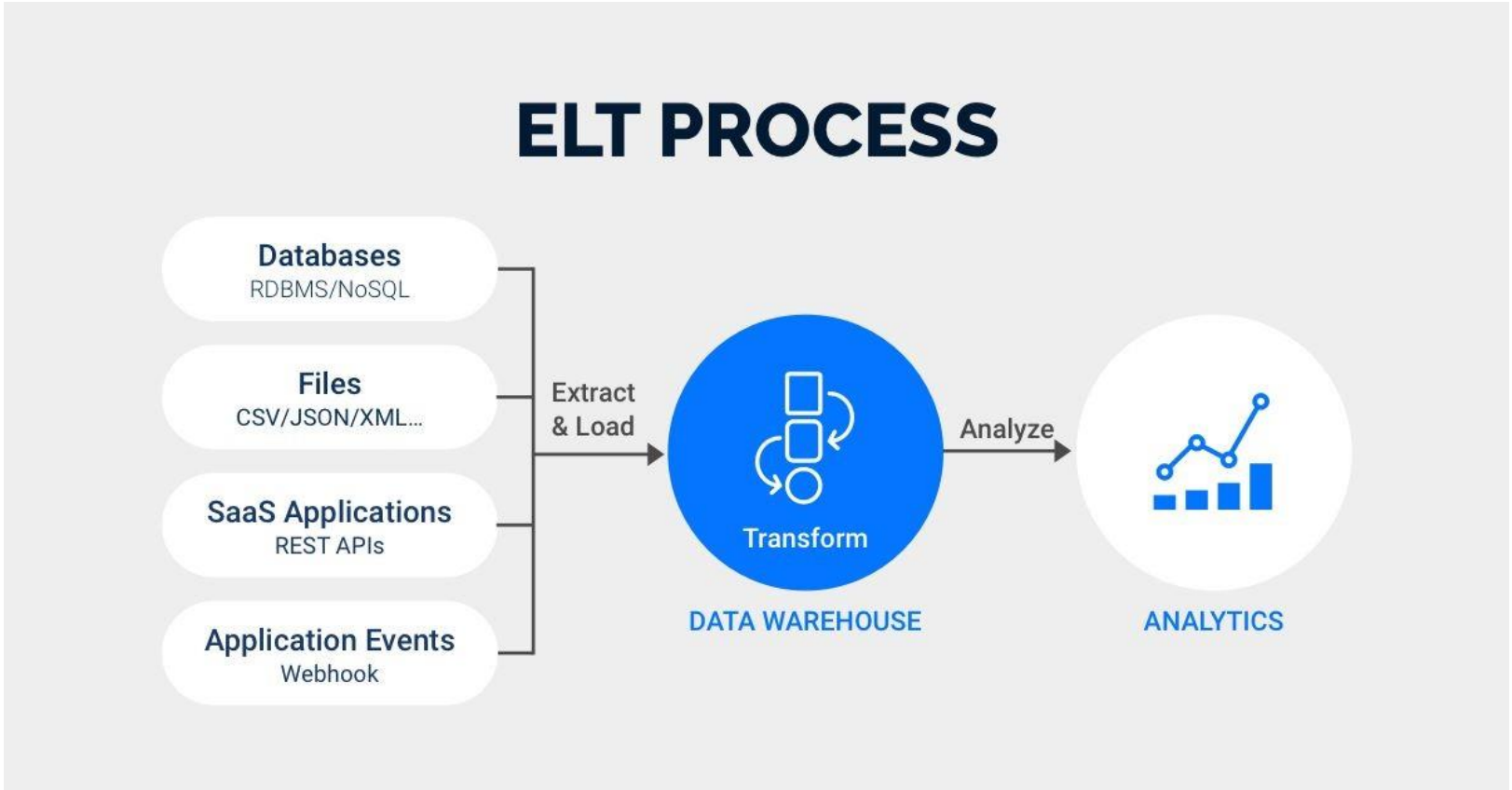
- Cost

Source: https://www.talend.com/resources/data-integration-methods/

ETL


ETL PROCESS
Databases
RDBMS/NoSQL
Files
CSV/JSON/XML...
SaaS Applications
REST APIs
Application Events
Webhook
Extract
Transform
STAGING AREA
Load
DATA WAREHOUSE
Analyze
ANALYTICS

ELT


ELT PROCESS
Databases
RDBMS/NoSQL
Files
CSV/JSON/XML...
SaaS Applications
REST APIs
Application Events
Webhook
Extract
& Load
Transform
DATA WAREHOUSE
Analyze
ANALYTICS

ETL vs ELT

Pros of ETL:

- **Compliance -** ETL is better for compliance with GDPR, HIPAA, CCPA and other standards
- **Implementation -** many ETL tools and experts
- **Maturity -** more documentation, tools and best practices

Pros of ELT:

- **Maintenance -** all data is always flowing with automated process
- **Speed -** data available in Data Warehouse faster since no transformation layer
- **Cost -** only cloud-based platforms needed at lower cost. On premise ETL processes require expensive hardware

Data Integration Tools

On-premise tools

- Oracle Data Service Integrator
- Informatica PowerCenter
- IBM InfoSphere Information Server

Cloud-based tools

- SnapLogic
- Talend Cloud Integration

Open-source tools

- Talend Open Studio
- Tibco Jaspersoft

Data Integration Best Practices

1. Identify your business needs first
2. Include internal business expert in the data integration team
3. Consider Long-Term Goals
4. Take into consideration the total cost of all methods
5. Avoid very complex data integration solutions
6. Choose a flexible solution

Module 8 - Document & Content Management

What is Document & Content Management

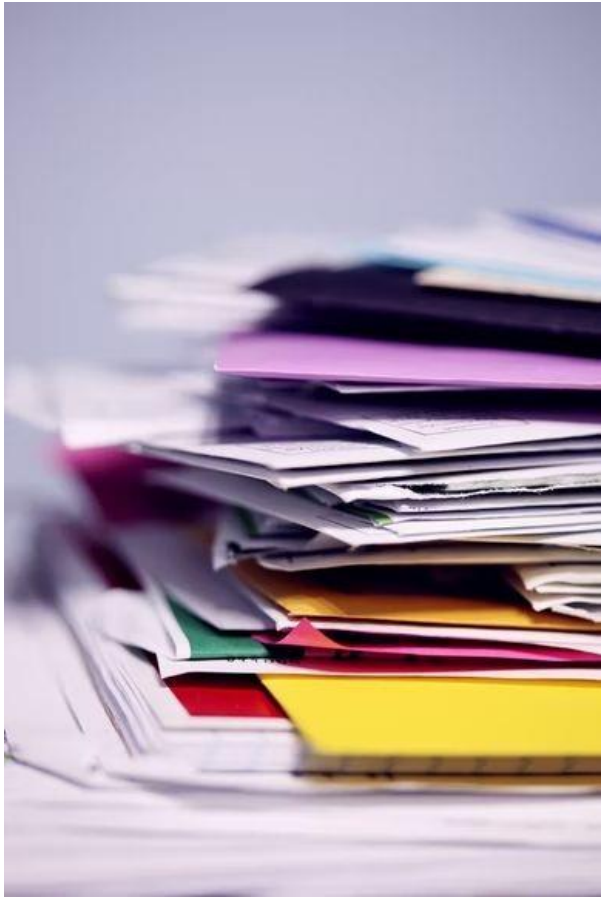
“**Document & Content management** is the process of establishing planning, implementation and control activities for lifecycle management of data and information found in any form or medium - outside of relational databases”

Why do we need Document & Content Management?

- Comply with legal obligations
- Comply with customer expectations regarding records management
- Effective and efficient storage, retrieval and use of documents and content
- Integration between structured and unstructured content

What is a DMS?

- DMS stands for Document Management System
- What are the benefits?
- Types of DMS



What is a CMS?

- CMS stands for Content Management System
- What are the benefits?
- Types of CMS

ECMS

- ECMS stands for Enterprise Content Management System
- What are the benefits?
- CMS vs ECM

Figure 1: Magic Quadrant for Content Services Platforms

Source: Gartner (October 2021)

Document Management vs Enterprise Content Management

Comparison	Document Management System (DMS)	Enterprise Content Management System (ECMS)
Type of Data	Structured data in traditional formats (Word, PDF, PowerPoint, Excel, etc)	Structured + unstructured data such as images, audito, video files, HTML, etc
Main purpose	Workflow management and regulatory compliance	Storage, retrieval and publishing of content
Key difference	DMS is a software	ECM is a set of tools and processes. ECM is a broader version of DMS
Company size	DMS only solution can work well for small companies	ECM solution needed in bigger organizations