Esoon Ko
IT Högskolan

Security And GDPR

Security

Security is a broad and critical concept that encompasses the measures and practices designed to protect systems, networks, and data from threats, vulnerabilities, and unauthorized access. In today's digital age, security is vital for safeguarding sensitive information, ensuring the privacy of individuals, and maintaining the integrity and availability of services.

Why Security in Data Engineering

Data Sensitivity: Data engineers handle a variety of sensitive data types, including personally identifiable information (PII), financial data, and intellectual property.

Regulatory Requirements: Compliance with laws and regulations such as GDPR, HIPAA, and CCPA is mandatory.

Business Impact: Data breaches can lead to financial loss, reputational damage, and legal penalties.

You are working with data: information that may be sensitive and are entrusted onto you to handle. It can be easy to forget about security when you focus on getting data to where it should.

CIA Triad

Confidentiality:

Ensures that sensitive information is accessible only to authorized individuals and entities.
Prevents unauthorized access, which could lead to data breaches or leaks.
Example: Encrypting data so that only those with the decryption key can read it.

Integrity:

Protects data from being altered or tampered with by unauthorized parties.
Ensures that data remains accurate, complete, and trustworthy.
Example: Using checksums or cryptographic hashes to detect changes in data.

Availability:

Ensures that systems and data are accessible to authorized users when needed.
Involves protecting systems from disruptions caused by attacks (e.g., DDoS), hardware failures, or natural disasters.
Example: Implementing redundancy and disaster recovery plans to maintain service continuity.

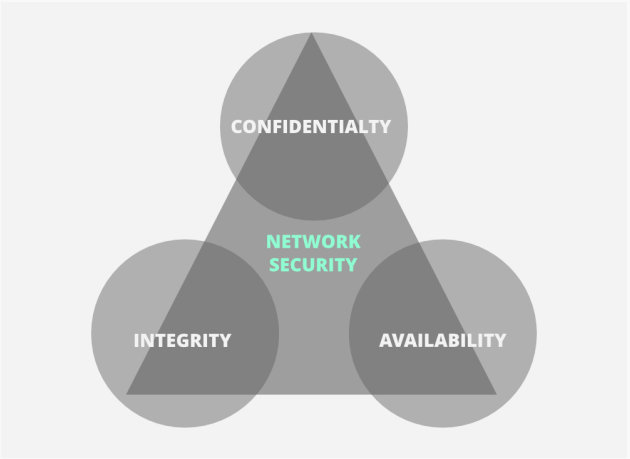


Image from GeeksforGeeks

Types of Security

- Physical Security
- Information Security
- Network Security
- Application Security
- Operational Security
- Cybersecurity

Types of Security

Physical Security:

- Protects the physical infrastructure (servers, data centers, and offices) from physical threats such as theft, vandalism, or natural disasters.
- Involves measures like surveillance cameras, security personnel, and controlled access to facilities.

Information Security:

- Focuses on protecting data in all forms—digital and physical—from unauthorized access, use, disclosure, disruption, modification, or destruction.
- Encompasses practices like encryption, access control, and data masking.

Network Security:

- Protects the integrity, confidentiality, and availability of data as it is transmitted across or accessed through networks.
- Involves firewalls, intrusion detection systems (IDS), and secure communication protocols (e.g., SSL/TLS).

Types of Security

Application Security:

- Secures software applications by identifying, fixing, and preventing security vulnerabilities in code.
- Practices include secure coding, code reviews, and the use of security frameworks.

Operational Security:

- Focuses on the processes and decisions that protect data and systems.
- Involves policies for password management, incident response, and security training for employees.

Cybersecurity:

- Specifically addresses threats that originate from cyberspace, such as hacking, malware, and phishing attacks.
- A subset of information security focused on protecting systems connected to the internet.

Attack Vectors

Attack vectors are the various paths or methods that malicious actors use to gain unauthorized access to systems, networks, or data. These vectors are essentially the entry points through which cyberattacks are carried out. By exploiting vulnerabilities in these paths, attackers can compromise security, steal data, disrupt operations, or cause other forms of harm.

Types of Attack Vectors

Malware:

Description: Malicious software designed to damage, disrupt, or gain unauthorized access to a system.

Examples: Viruses, worms, ransomware, spyware, trojans.

How it Works: Malware can be delivered via email attachments, malicious websites, or infected software downloads. Once installed, it can perform various harmful activities, such as stealing data, encrypting files for ransom, or taking control of a system.

Phishing:

Description: A social engineering attack where attackers trick individuals into revealing sensitive information by pretending to be a trustworthy entity.

Examples: Fake emails from banks asking for login details, fraudulent websites mimicking legitimate services.

How it Works: The attacker sends a deceptive message (often via email) that appears legitimate, urging the recipient to click a link, download an attachment, or provide personal information. Once the victim complies, the attacker can gain access to their accounts or inject malware into their system.

Man-in-the-Middle (MITM) Attacks:

Description: An attack where the attacker secretly intercepts and possibly alters communication between two parties.

Examples: Eavesdropping on unencrypted Wi-Fi traffic, intercepting communications on a compromised network.

How it Works: The attacker positions themselves between the victim and the intended destination (e.g., a website), intercepting the data exchanged. They can then read, alter, or inject malicious content into the communication without the victim’s knowledge.

Types of Attack Vectors

SQL Injection:

Description: A code injection technique that allows attackers to execute malicious SQL queries against a database.

How it Works: Attackers exploit vulnerabilities in web applications by injecting malicious SQL code into input fields (like login forms). If the application does not properly validate input, the injected code can be executed, allowing attackers to access, modify, or delete database data.

Cross-Site Scripting (XSS):

Description: An attack where malicious scripts are injected into trusted websites.

How it Works: The attacker injects a script into a web page that other users visit. When a user visits the page, the script is executed in their browser, potentially stealing cookies, session tokens, or redirecting the user to malicious sites.

Denial of Service (DoS) and Distributed Denial of Service (DDoS):

Description: Attacks aimed at making a system, service, or network unavailable to its intended users by overwhelming it with traffic.

How it Works: In a DoS attack, the attacker uses a single system to flood a target with traffic. In a DDoS attack, the attacker uses multiple systems (often part of a botnet) to launch a coordinated attack, overwhelming the target and causing it to slow down or crash.

Types of Attack Vectors

Insider Threats:

Description: Threats that originate from within an organization, often involving individuals with authorized access to systems and data.

Examples: Employees stealing sensitive data, accidentally leaking confidential information, or deliberately sabotaging systems.

How it Works: Insiders may misuse their access privileges for personal gain, to damage the organization, or unintentionally cause a security breach due to negligence or lack of awareness.

Drive-By Downloads:

Description: Attacks where malware is automatically downloaded to a user's device without their knowledge or consent.

How it Works: Attackers exploit vulnerabilities in web browsers or plugins. When a user visits a compromised or malicious website, malware is downloaded and executed on their device without any user interaction.

Password Attacks:

Description: Attacks aimed at cracking or stealing user passwords.

Examples: Brute force attacks, dictionary attacks, keyloggers, credential stuffing.

How it Works: Attackers use various methods to guess or steal passwords, such as trying multiple combinations (brute force), using common passwords (dictionary attack), or capturing keystrokes (keyloggers). Once they obtain a password, they can access the victim's accounts or systems.

Types of Attack Vectors

Social Engineering:

Description: Manipulating individuals into divulging confidential information or performing actions that compromise security.

Examples: Phishing, pretexting, baiting, tailgating.

How it Works: Attackers exploit human psychology to trick individuals into providing sensitive information or bypassing security measures. This could involve posing as a trusted authority, creating a sense of urgency, or appealing to curiosity.

Mitigating Attack Vectors

- **Regular Security Updates:** Ensure all systems, software, and applications are up-to-date with the latest security patches.

- **Strong Authentication:** Implement multi-factor authentication (MFA) to add an extra layer of security.

- **User Education:** Train employees and users to recognize and avoid phishing, social engineering, and other common attacks.

- **Network Security:** Use firewalls, intrusion detection systems (IDS), and secure communication protocols to protect networks.

- **Input Validation:** Ensure that all user inputs are validated and sanitized to prevent SQL injection and XSS attacks.

- **Encryption:** Encrypt sensitive data both in transit and at rest to protect it from interception or unauthorized access.

Access Controls: Implement the principle of least privilege, ensuring that users have the minimum necessary access to perform their roles.

Mitigating Attack Vectors and Best Practices

- **Regular Security Updates:** Ensure all systems, software, and applications are up-to-date with the latest security patches.

- **Strong Authentication:** Implement multi-factor authentication (MFA) to add an extra layer of security. Work with Identity and Access Management.

- **User Education:** Train employees and users to recognize and avoid phishing, social engineering, and other common attacks.

- **Network Security:** Use firewalls, intrusion detection systems (IDS), and secure communication protocols to protect networks. Constantly work with monitoring and logging for suspicious activities.

- **Input Validation:** Ensure that all user inputs are validated and sanitized to prevent SQL injection and XSS attacks.

- **Encryption:** Encrypt sensitive data both in transit and at rest to protect it from interception or unauthorized access.

Access Controls: Implement the principle of least privilege, ensuring that users have the minimum necessary access to perform their roles. Use data masking to hide sensitive data.

Data Best Practices

Data Encryption:

At Rest: Use encryption methods like AES-256 for databases and storage systems.

In Transit: Implement SSL/TLS to secure data traveling over networks.

Access Controls and Authentication:

Role-Based Access Control (RBAC): Restrict access based on the user's role within the organization.

Multi-Factor Authentication (MFA): Combine something the user knows (password) with something they have (a token) or something they are (biometric verification).

Data Masking and Anonymization:

Masking: Hide sensitive data fields when sharing data sets with non-authorized users.

Anonymization: Remove or modify personal identifiers to prevent re-identification of individuals.

Secure Data Storage:

Secure Cloud Storage: Use services like AWS S3 or Cloud Storage with encryption and strong access controls.

Database Security: Regularly update database software, use encrypted backups, and apply patches.

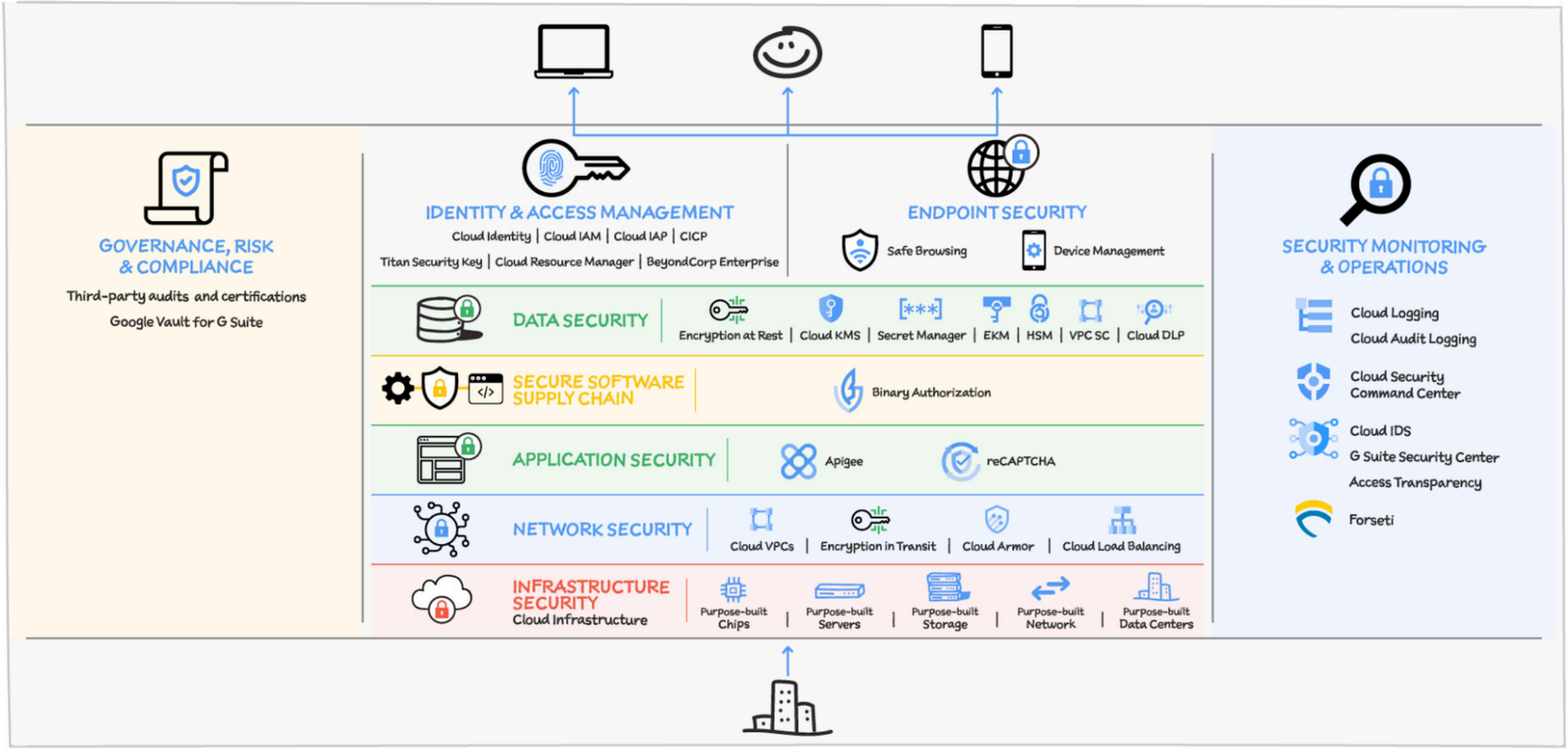
Monitoring and Logging:

Real-Time Monitoring: Tools like Splunk or ELK stack can help monitor for unusual activities.

Log Retention Policies: Keep logs for an appropriate duration to detect and investigate incidents.

Google Cloud Security Controls

#GCPSketchnote @PVERGADIA THECLOUDGIRL.DEV 11.07.2021



Discussions

Data Encryption:
You have a python script in **Google Cloud Run Functions**

What **potential vulnerabilities** are there?

Discussions

Data Encryption:
You have data in **Google Bigquery**

What **potential vulnerabilities** are there?

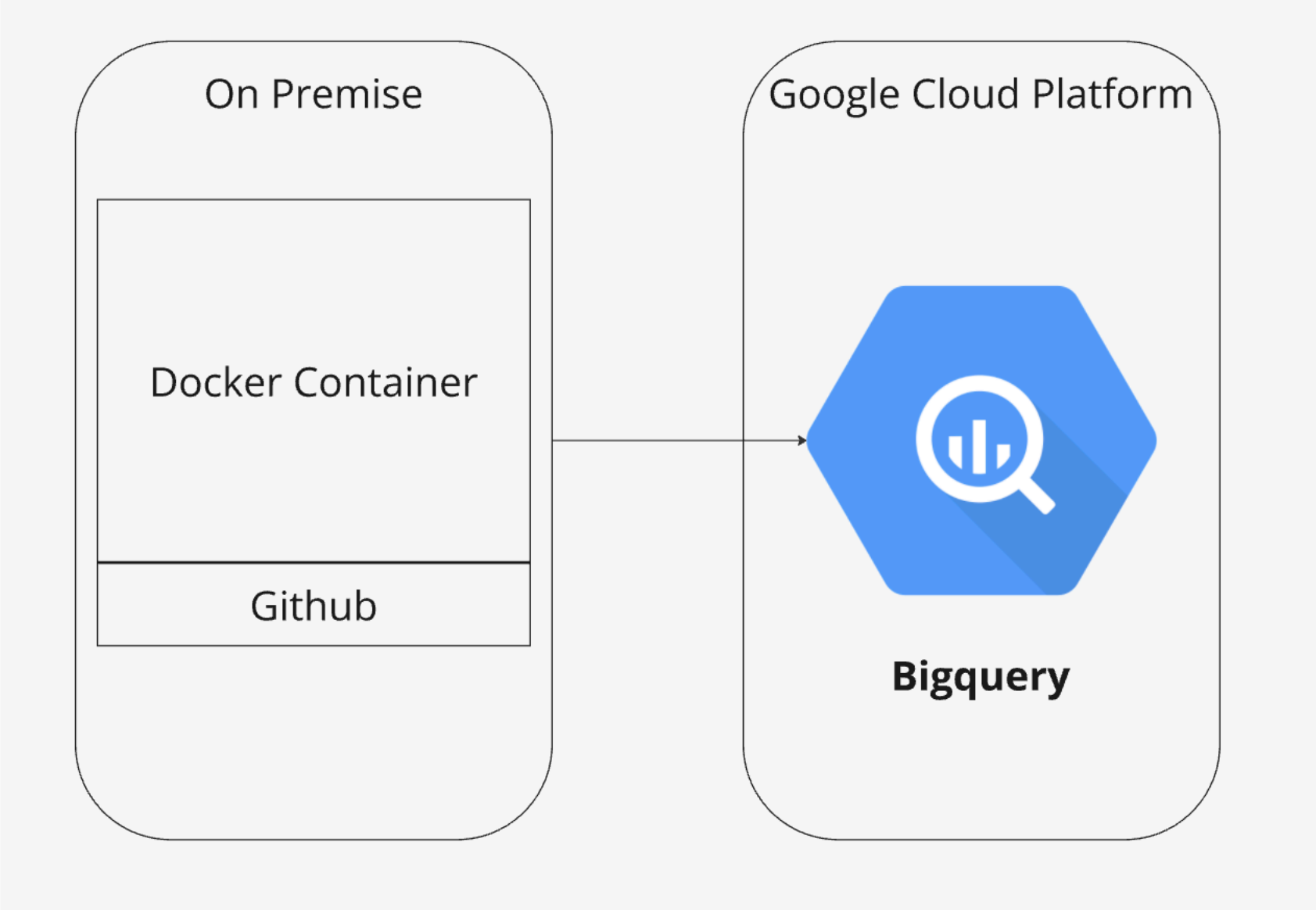
Discussions

Data Encryption:

You have the following setup.

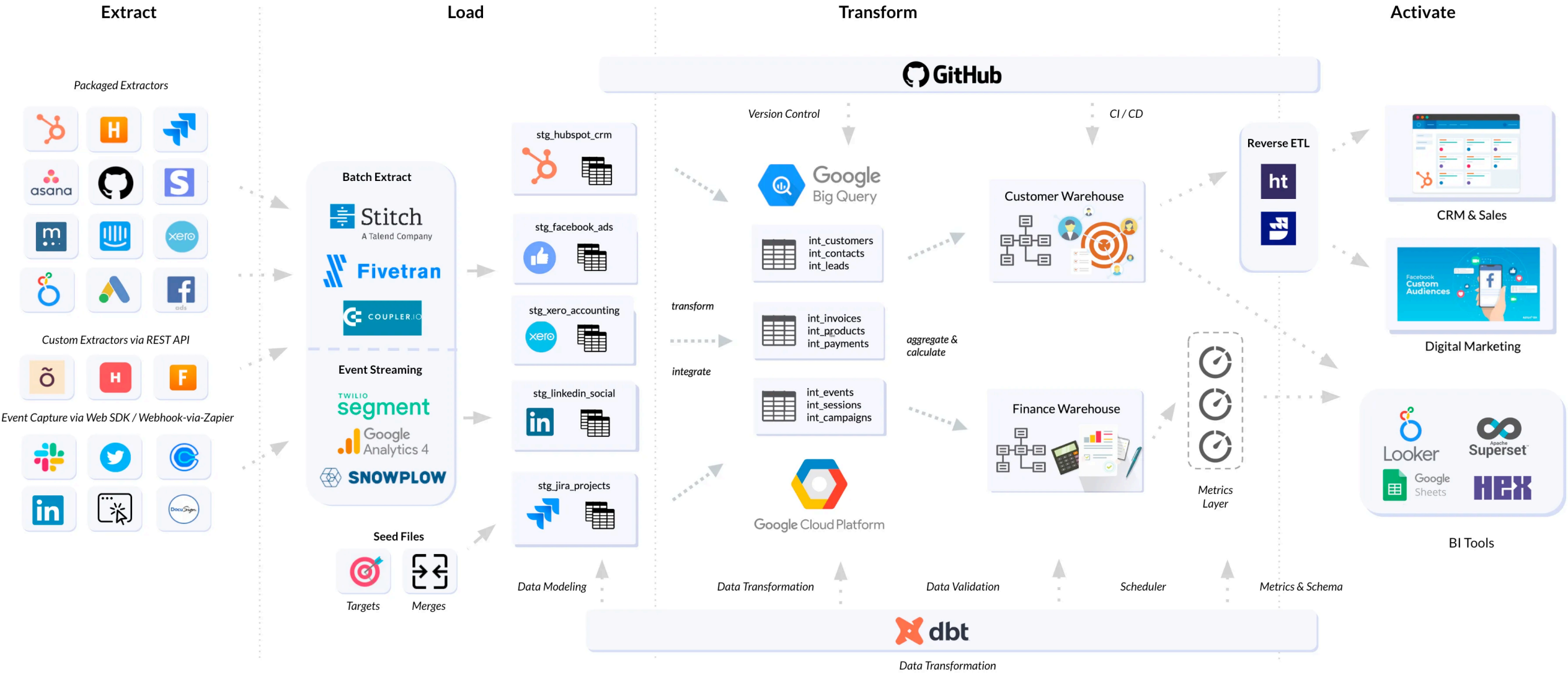
You have a docker container on premise that sends data up to Bigquery in Google Cloud Platform.

What **potential vulnerabilities** are there?



Discussions

How about here? Lets look at the components - From Rittman Analytics



GDPR - General Data Protection Regulation

The **General Data Protection Regulation (GDPR)** is a comprehensive data protection law that was enacted by the European Union (EU) and came into force on May 25, 2018. GDPR is designed to give individuals more control over their personal data, streamline regulations across Europe, and impose strict rules on how organizations handle personal data.

Objectives: Strengthen and unify data protection for individuals within the EU, and give control back to citizens over their personal data.

Complying to GDPR - Who and what?

Global Reach:
GDPR applies to any organization processing personal data of EU citizens, regardless of where the organization is located. This includes companies based outside the EU that offer goods or services to, or monitor the behavior of, EU residents.

Material Scope:
GDPR applies to the processing of personal data by automated means (e.g., data stored in electronic systems) and to manual filing systems where personal data is accessible according to specific criteria.

Penalties:
Non-compliance can result in fines up to €20 million or 4% of global turnover, whichever is higher.

Obligations:
Communication to Authority: Data controllers have the obligation to notify the relevant supervisory authority of a data breach within 72 hours of becoming aware of it, unless the breach is unlikely to result in a risk to individuals' rights and freedoms.

Communication to Data Subjects:
If the breach is likely to result in a high risk to the rights and freedoms of individuals, the data controller must also inform the affected individuals without undue delay.

Documentation:
Data controllers must document all data breaches, including the facts relating to the breach, its effects, and the remedial action taken.

Key GDPR Concepts Relevant to Data Engineering

Personal Data:

Definition: Any information relating to an identified or identifiable person (data subject).
Examples: Names, email addresses, IP addresses, cookies, and even indirect identifiers.

Data Subject Rights:

Right to Access: Individuals can request a copy of their data from a controller.
Right to Rectification: Data subjects can request corrections to inaccurate or incomplete data.
Right to Erasure (Right to be Forgotten): Individuals can request deletion of their data under certain circumstances.
Right to Data Portability: Enables data subjects to transfer their data between controllers.

Data Minimization and Purpose Limitation:

Minimization: Only collect data that is necessary for a specific purpose.

Lawful Bases for Processing:

Consent: Obtaining explicit permission from data subjects.
Legitimate Interest: Processing that is necessary for the legitimate interests of the data controller.

Data Protection by Design and by Default:

By Design: Integrate data protection measures into the development of business processes and IT systems.
By Default: Ensure that data protection is the default setting for any system.

Data Breach Notification:

Requirement: Notify the relevant supervisory authority within 72 hours of becoming aware of a data breach.
Communication: If the breach is likely to result in a high risk to the rights and freedoms of individuals, they must also be informed.