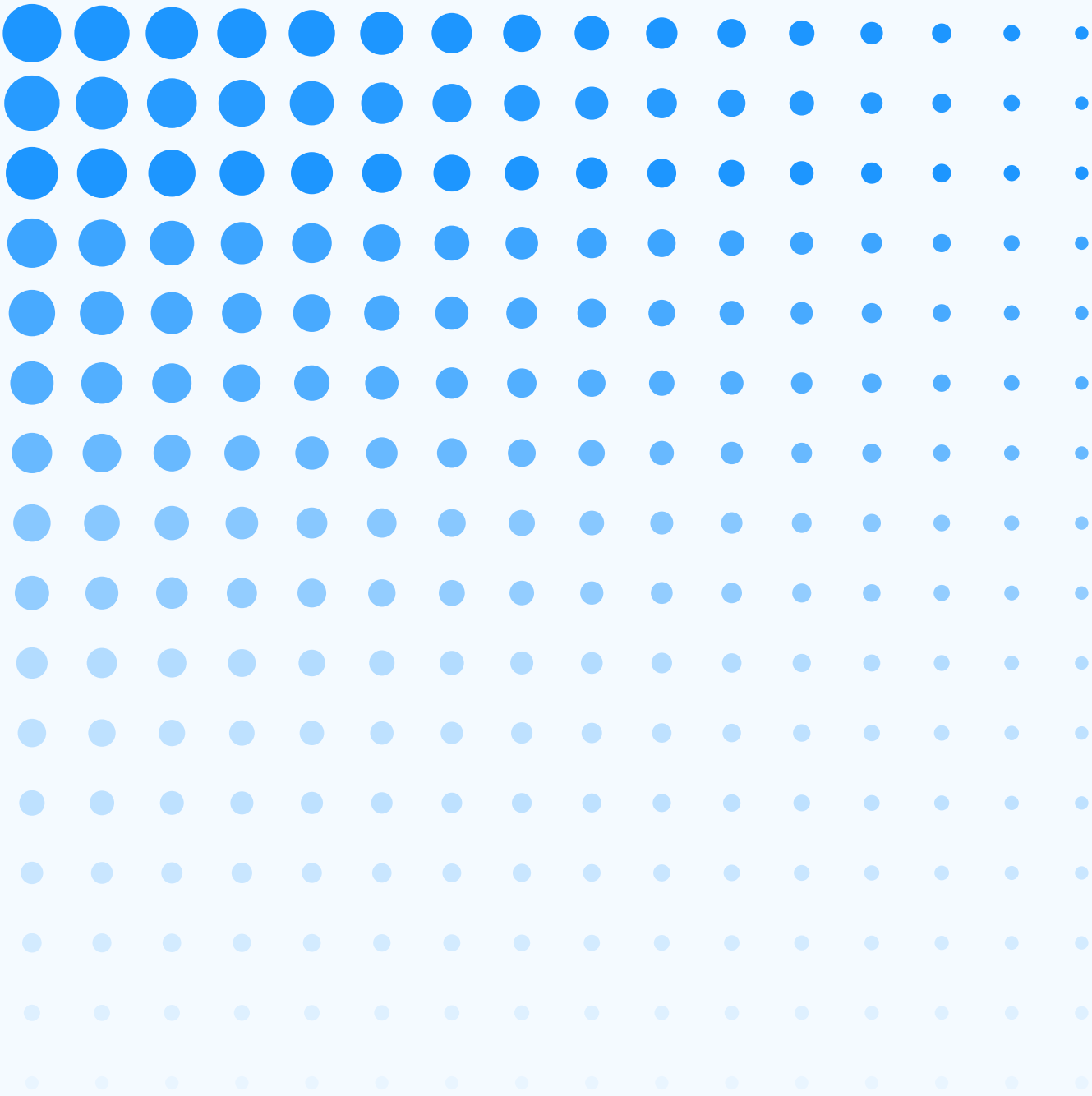
The Complete Guide to Data Security on Heroku

PROTECTION FOR YOUR DATA



integrate.io

Table Of Contents

01

INTRODUCTION

Keeping Heroku Data Safe

Heroku is a powerful Platform-as-a-Service (PaaS) that enables all kinds of business applications. Most importantly, it's part of the Salesforce family, which means you can integrate it with your sales and marketing functions.

Given the importance of Heroku and Salesforce, you need to take extra care when authorizing transactions on the platform. The slightest slip-up could expose personally identifiable information (PII) and possibly lead to a data breach.

Where Is Heroku Vulnerable?

Each Heroku product has different vulnerabilities:

• Heroku Platform: Unauthorized user access or malfunctioning apps may expose data

• Heroku Postgres: SQL injections or misconfigured connectors might allow direct database access

• Heroku Connect: The direct connection between Heroku and Salesforce is a rich target for hackers

• Heroku Elements: Marketplace add-ons may create new vulnerabilitiesWhile Heroku is generally safe, it's important to know the risks and act to protect your data.

The Risks of a Data Breach

Cybercriminals never miss an opportunity. In 2019, one gang used Heroku's flexibility to build a highly effective skimming app that stole credit card numbers[1]. If they find a way to exploit your data structure's vulnerabilities, they will take it.

Data breaches can be highly damaging and costly to repair. IBM estimates an average recovery cost of $3.86 million[2], although this varies according to the severity of the breach. That's on top of the reputational damage associated with a data breach, which can lead to loss of business in the long run.

The cost of data breaches

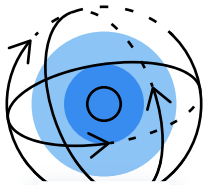
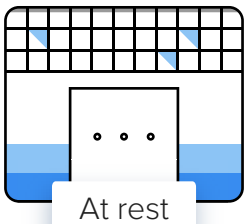
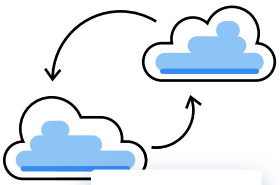
A data breach is any exposure of sensitive data to an unauthorized party. Such parties can include hackers, rogue employees or any person who isn't authorized to access the data.

Dealing with a data breach is expensive. IBM estimates the average recovery cost at $3.86 million. This figure includes immediate restorative action, as well as loss of business and reputational damage that arises from losing customer data. Regulatory fines can push this figure even higher. In 2017, Equifax received a world-record fine of $575 million for a data breach that resulted from a missed Apache update.

Data breaches can also cause real human suffering. Over 650,000 people experienced identity theft in 2019, and many of these cases were directly linked to data breaches. When a customer provides you with their private data, they're trusting you to keep them safe.

When is data most at risk?

Data can be in one of three states: in use, at rest or in transit. Each of these states has its own level of risk.

What it means	State	What's the risk
The data is in the memory of a production system. This could be an automated system, such as an eCommerce module, or a user-facing system like the CMS.	 In use	Unauthorized users may be able to capture this data. If hackers can obtain credentials, they could log in and steal information.
The data is stored in a data repository, such as a data warehouse. This may be on-premise or in the cloud.	 At rest	Unauthorized parties might be able to access unencrypted data repositories and backups. Data warehouses need to be integrated with other systems, or crucial data may not be available when needed.
The data is moving from one location to another. This could be synchronization between system, or a user accessing a remote system. Transit can be in-premise or across the internet.	 In transit	Hackers can intercept unencrypted data as it moves between locations. If a user is on public wi-fi, they may be especially vulnerable to attack.

The basics of data security

In this guide, we'll look at how to develop an effective data security strategy. The key elements to bear in mind are to:

1. Management buy-in to support a security-first culture.
2. Create a data security policy.
3. Understand your compliance requirements.
4. Categorize your data according to sensitivity.
5. Control access to data.
6. Use a trusted ETL partner to protect data in transit and during transformation.

If you'd like to learn more about how Integrate.io can protect data in transit, you can jump ahead to chapter six.

02

DEVELOPING YOUR DATA SECURITY POLICY

Data security is a philosophy. To protect against breaches, you have to make data security part of your core values and live those values every day. And for that, you require a Data Security Policy (DSP) that everyone can access.

Data security policies help to balance the three main elements of security:

C — **Confidentiality:** Sensitive information must be safe from prying eyes.

I — **Integrity:** Data must be free from corruption or loss.

A — **Availability:** Data must always be available for legitimate business purposes.

These elements, known as the CIA triad, are sometimes in competition with each other, but a strong DSP will help you find balance.

How to Write a Data Security Policy

Data security policy (or DSP) documents are often written in dense legal language, usually in the hope that they might help limit the organization's liability in the event of a breach. However, ignorance or a lack of understanding of your company's data security policy doesn't constitute a legal defense in the event of a breach.

Instead, it's usually better to write a DSP in plain language that most people can understand. Doing so means that the DSP document acts as a useful guide for employees, customers and partners, helping them understand how to keep data safe.

Your DSP document will generally include the following sections:

- **Purpose:** Your main goals, which helps readers understand the spirit of the law. Examples of such goals are: protect customer confidentiality; safeguard business reputation; comply with relevant laws; proactively work to minimize the risk of breaches.
- **Scope:** What does the DSP cover? The section will explicitly scope in sensitive data such as PII and classified information. It's also useful to scope out certain things, such as publicly available information and non-sensitive data.
- **Data classification:** An overview of how you classify data, ranging from Public to Highly Sensitive. We'll look at this in more depth in chapter 4.
- **Policies:** A list of fundamental policies that relate to all data activity on the network. This will include things like:
    - **Acceptable Usage Policy (AUP):** Terms and conditions for people using your network, such as employees or customers, who may need to sign a copy of an AUP agreement.
    - **Data in transit rules:** Standards for data transportation, including minimum levels of data encryption.
    - **Physical security policy:** Standards for keeping physical devices safe, which includes entry to the office building as well as transportation of electronic devices.
    - **Access Control Policy:** Rules about who may access data and how.
    - **Cloud adoption policy:** Guidelines on implementing cloud-based systems, including data warehouses and ETL.
    - **Remote access policy:** Outline the requirements for external access to local systems.
    - **Change management policy:** The process for deploying new systems or procedures, such as minimum documentation requirements.
    - **Business continuity and disaster response plans:** A plan for what happens in the event of a disaster or catastrophic failure.
- **Technical guidelines:** The DSP may outline specific technical standards for the organization. This can include things such as:
    - Operating systems.
    - Database management systems.
    - Encryption standards.
    - User authentication protocols.
    - Remote access tools.
    - Analytics and event logging.
- **Reporting and oversight:** The DSP will outline an internal audit process to ensure that all systems meet the agreed requirements. Any issues will go back to a specified authority, such as the data governance team.
- **Update procedure:** A Data Security Policy is a living document that grows with the organization. You'll need an established process to review and update the DSP to keep up with new technology and stay abreast of new threats.

Data Security Policy checklist

A data security policy can't cover every eventuality. Instead, the goal is to create a framework that offers guidance when someone needs to make a decision about your organization's data. Here is a checklist of questions to help establish if your document meets standards.

Have you clarified the DSP with all major stakeholders, including your executive team, I.T., H.R. and compliance?

Have you clarified rules about permissions and access roles? (see chapter 5)

Is the document written in clear language?

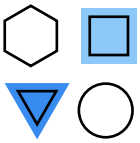
Is there a framework for working with trusted partners? (see chapter 6)

Have you outlined your primary data security goals?

Is there a clear Acceptable Usage Policy for general users?

Do your policies accord with all compliance requirements? (see chapter 3)

Have you established a way to ensure that everyone follows DSP guidelines?



Have you categorized data according to sensitivity? (see chapter 4)

Is it easy to update the document as required?

Once you have answered these questions, you're ready to publish your security policy. It's good practice to try to engage people in conversation about the policy, using channels such as eLearning tools or discussion seminars.

03

UNDERSTANDING DATA SECURITY COMPLIANCE LAWS

Data security practices are closely related to the legal concept of data protection. Under data protection rules, organizations have an obligation to protect individual confidentiality. This means that you have to keep data safe, prevent unauthorized access and only use data for legitimate purposes.

Data protection laws vary across countries and even between states. However, many laws have an extra-territorial effect, which means that authorities will punish foreign companies for breaches.

Main data security compliance laws

Here are a few of the primary laws you need to be aware of:

1. General Data Protection Regulation (GDPR)
2. Bundesdatenschutzgesetz (BDSG)
3. Health Insurance Portability and Accountability Act (HIPAA)
4. California Consumer Privacy Act (CCPA)
5. Australian Privacy Act of 1988
6. Lei Geral de Proteção de Dados (LGPD)

General Data Protection Regulation (GDPR)

Primary jurisdiction: European Union
Data covered: Any data that could potentially identify an E.U. citizen
Website: https://gdpr-info.eu/
Notes: GDPR is one of the most stringent data protection regimes in the world. Companies must allow users to opt out of data collection, and they can only capture PII for essential business purposes. Organizations face severe restrictions on transporting PII out of Europe, even when using a third party service. The E.U. has successfully fined a number of American firms for GDPR breaches, including Google [5].

Bundesdatenschutzgesetz (BDSG)

Primary jurisdiction: Germany
Data covered: Any data that could potentially identify a German citizen
Website: https://www.gesetze-im-internet.de/englisch_bdsg/index.html
Notes: E.U. member states can introduce their own laws to supplement GDPR. Germany is the only state to have done so to date, with the BDSG law that imposes stricter controls and steeper fines. German citizens can claim for non-monetary damages such as stress and suffering under BDSG.

Health Insurance Portability and Accountability Act (HIPAA)

Primary jurisdiction: United States
Data covered: Protected Health Information of Americans
Website: https://www.hhs.gov/hipaa/
Notes: HIPAA refers specifically to health information about an individual, which includes medical records and biometric information. Under HIPAA, data handlers must ensure confidentiality, integrity and availability of all relevant information. They must also take steps to prevent breaches and unauthorized access.

California Consumer Privacy Act (CCPA)

Primary jurisdiction: California
Data covered: Personal Identifiable Information (PII) of Californian consumers
Website: https://oag.ca.gov/privacy/ccpa
Notes: CCPA grants consumers more power over their PII, including the right to know what's on file, the right to request deletion and the right to opt out of the sale of PII. In the event of a compliance breach, consumers can directly sue the company. This law is currently unique in the U.S., but it is the template for forthcoming legislation in other states [6].

Australian Privacy Act of 1988

Primary jurisdiction: Australia
Data covered: PII of Australian citizens
Website: https://www.ag.gov.au/rights-and-protections/privacy
Notes: Australia amended its 1988 Privacy Act in 2017 to cover digital communications. The act takes a principles-based approach to compliance, so companies have some freedom as long as they follow the spirit of the principles. Since 2018, companies have been obliged under the Privacy Act to notify Australian authorities of data breaches that may cause harm to an individual.

Lei Geral de Proteção de Dados (LGPD)

Primary jurisdiction: Brazil
Data covered: Any data that could potentially identify a Brazilian citizen
Website: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm
Notes: Brazil's LGPD is one of the first international law to model itself on the E.U.'s GDPR. As with European law, the LGPD covers a wide range of personal information and has an extra-territorial effect on foreign companies. However, LGPD is generally less punitive in terms of fines and enforcement.

04

CLASSIFYING DATA BY SENSITIVITY

Data classification hinges on one question: What would be the consequences if this data leaked?

If you consider this question in terms of your company's data, you see three main categories:

- **High Impact**
  This category includes personal information that could breach data protection laws or expose people to the risk of identity fraud. It also includes sensitive corporate documents such as confidential reports and strategy documents.
- **Moderate Impact**
  This includes information that you'd rather keep private, but which poses no immediate risk. For example, B2B invoices and supplier agreements may fall into this category, as well as personal information that doesn't identify an individual.
- **Low Impact**
  This information won't hurt your business if leaked, and much of it may already be available to the public. Press releases, white papers, and non-proprietary corporate information all fall into this category.

Companies can use this system to create a classification taxonomy for data. Some of the more common systems use Public, Internal, Confidential, and Restricted. You can create further compartments within these general accepted categories as well.

How to classify personal data

Most privacy laws define PII as information that could potentially reveal someone's identity. Clearly, this includes unique identifiers such as:

- **Name**
- **Address**
- **Date of birth**
- **Login credentials**
- **Social security**
- **IP address**
- **Biometric information**

It doesn't mean that all records associated with an individual automatically count as PII. For example, a register of login times for a user account is personal information, but it is not necessarily identifiable.

That said, data owners must bear in mind that minor pieces of data can reveal someone's identity when combined. A study by Sophos found that a combination of gender, date of birth and ZIP code is enough to uniquely identify 87 percent of U.S. residents[7].

When in doubt, it's best to assume that all personal records count as PII until you're sure otherwise.

Expanding your data classifications

The system above describes an outcome-based data classification system. Some organizations may choose to add extra layers of detail to create a more expressive taxonomy that describes multiple types of risk.

Some of the extra factors to consider are:

- **Frequency of movement**
  Data is at risk when it keeps moving between locations. Conversely, the risk decreases when the data remains encrypted in a secure repository and rarely moves.
- **Encryption and password protection**
  Additional measures can help lower risk, such as password protecting files or encrypting them in transit. It's not always possible to encrypt in-use data, so this increases the potential risk.
- **Access level**
  The more people with access, the greater the risk. If data rests in a highly restrictive environment, it's low risk. Data on a live system with multiple users is at a much higher risk.
- **Compliance impact**
  Some organizations choose to classify data according to legislation. Health data poses a high risk of HIPAA breaches, while E.U. data could lead to a GDPR issue.

Classifying data helps to support data security while also improving performance. If you arrive at a set of definitions that meets your business needs, you can make sure that highly sensitive data always has the best possible protection. Then you can focus on improving processing efficiency for low-risk data.

05

BUILDING A SECURITY STRATEGY ON IDENTITY

Users tend to be the weakest point of any network. Cybercriminals often focus on stealing login credentials through phishing or social engineering attacks. When they succeed, they can virtually walk through the front door unchallenged. Rogue or poorly trained employees may also pose an internal threat to data security.

This threat is greater in an age of remote working and BYOD policies. To meet this challenge, we have to change the way of thinking about user security. Instead of being a weakness, user identity can be the primary perimeter for data security.

The Identity Perimeter

- Users have a single login with strong authentication, including multiple factors
- Analytics-powered systems monitor each identity for unusual activity
- All data activity is tied to a specific user identify
- Role Based Access Control (RBAC)
- Compromised identities are easy to deactivate or delete

Key elements of an identity-based strategy

An identity strategy must respect all three elements of the CIA triad:

- **Confidentiality**

  User data is never exposed to an unauthorized party. Equally, users cannot access another person's data unless they have a good reason.
- **Integrity**

  Users must have access to quality data and be able to amend as required. Users should not be able to make any unauthorized changes or deletions.
- **Availability**

  Users should have access to the data they require with as few barriers as possible.

Identity strategy is about finding a balance between these elements. If your measures are too complex and restrictive, users won't be able to work with your system, which may force them to adopt risky workarounds. If your measures are too lax, you risk exposing sensitive data.

In practice, it's often a matter of implementing a system and then fine-tuning it according to your business needs. Follow the best practices below to find a balance that works.

Identity strategy best practices

- **Offer single sign-on**

  Multiple logins can pose a risk. Users tend to forget multiple passwords, so they end up reusing logins or, even worse, keeping a list of passwords on their desk. Single sign-on means that one username, one password, and one user identity linked to all data activity.

- **Set strong password standards**

  Password best practice techniques include:

  - At least 12 characters
  - Variety of ASCII characters, including letters, numbers, and symbols
  - Forbid the use of names, dictionary words or dates
  - Check against a password dictionary to see if a password is commonly used
  - Use randomly generated passwords where possible

- **Use two-factor authentication (2FA)**

  2FA is essential in a single sign-on environment. By requiring two forms of authentication from two different sources, it adds an extra step that most hackers can't replicate, even if they obtain a username and password. Typically you will use two of the following "factors" for your authentication:

  - Something you know (e.g. a password)
  - Something your have (e.g. phone, keys, etc)
  - Something you are (e.g. biometrics, fingerprints, etc)

- **Implement role-based access control**

  Each individual identity should be linked to a role, such as sales, analytics, customer service, etc. You can then implement a data access policy that reflects each role's needs, giving everyone the data they require. In the event of an organizational change, you can then change the configuration for a role rather than updating individual users.

- **Take a "least privilege" or "need to know" approach**

  Each role should have access to the data they need, but nothing else. This approach limits the chances of unauthorized data access by a rogue user, or via stolen login credentials.

- **Communicate with your users**

  In the perfect identity-based strategy, users will have a seamless data experience, while unauthorized parties will find it impossible. It's only possible to reach this level by working with your users and ensuring that they have the right level of data availability and an easy login process.

Education plays a vital role in this strategy. Ensure that everyone receives adequate training and support to understand how identity management plays a role in organizational security. Let them know what they need to do to keep data safe.

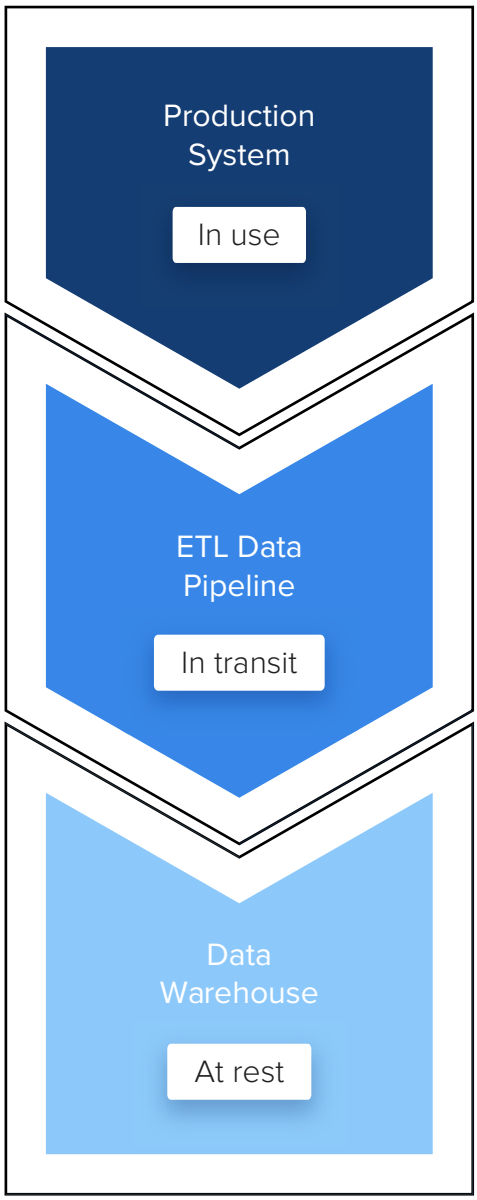
06

WORKING WITH A TRUSTED ETL PARTNER

Extract, Transform, Load (ETL) is a core process that allows you to store data in a secure repository. The process goes like this:

- **Extract:** Obtain data from live production systems, such as CMS, ERP, eCommerce, marketing automation and so on.
- **Transform:** Integrate and transform the raw data so it's suitable for storage in a data repository.
- **Load:** Send the transformed data to a secure storage location, such as a data warehouse.

During this process, the data passes through each of the three data states, as shown here



Data in transit can be vulnerable, especially when it is moving outside of the on-premise data environment, so must be encrypted. As most organizations are now reliant on cloud-based warehouses, this kind of data movement is an inevitable fact of life.

The main approaches to ETL are to build your own solution, install an off-the-shelf ETL locally or use a cloud ETL service.

	What is it?	Pros	Cons
In-house development	Internal dev team creates a bespoke ETL for your specific needs	Full control and transparency of all aspects of ETL	You need an in-house team to build the solution and provide ongoing support
Local ETL Install	Purchase an ETL solution and install it on your on-premise infrastructure	Control over configuration without needing to develop the software from scratch	Difficult to upscale and may not integrate securely with cloud-based warehouses
Cloud ETL	A third-party service manages your ETL needs across the cloud	Simple, no-code integration with cloud and on-premise services, with a trusted partner guaranteeing security	Works best with other cloud services, such as AWS, Salesforce, and cloud-based analytics tools

How do Cloud ETL providers guarantee data security?

Cloud-based ETL provides your data with a single point of egress from the network. Rather than having multiple pipelines connecting each production database to a repository, each production system has a secure connection to the ETL service. The ETL then has a separate connection to the data repository.

Data makes a pitstop on the ETL servers, where it passes through the transformation layer. With in-house or local solutions, this stage can be vulnerable.

However, a good cloud ETL provider such as Integrate.io takes significant steps towards protecting data security. This includes things like:

- **Security-first development process**
  Any reputable ETL vendor will start with security protocols before they even begin developing the service. As a user, you can identify trustworthy products by examining their security details and seeing if the provider has baked security into their product.
- **Physical security**
  The physical location of your data in transition is a major factor in security. Reputable vendors will guarantee that all servers are safe. Integrate.io uses an AWS infrastructure, which sits on Amazon's SOC-compliant data centers.
- **Working with reputable vendors**
  ETL is an interactive product by nature, with automatic integration to other services. ETL vendors must carefully vet and monitor all of their partners to ensure that their customers are not exposed to risk. The ETL vendor also stays on top of changing API requirements to ensure that integrations always meet current requirements.
- **Plans for disaster recovery**
  ETL vendors provide a critical connection between live systems and repositories. Real-time analytics services are dependent on an uninterrupted flow of data between these points. The ETL vendor should have a robust plan for maintaining service in any circumstances.
- **SOC 2 compliance**
  Routine testing is a requirement of this security standard. Integrate.io, for example, undergoes third-party penetration testing each year. When choosing an ETL vendor, ask to see their SOC 2 report & PenTest results before signing up.

07

ESSENTIAL CLOUD ETL DATA SECURITY FEATURES

There are many cloud-based ETL providers on the market, each offering a range of attractive features. For example, Integrate.io offers benefits like:

- High-speed transformations on a staging server
- Automated integration with most major production systems and data repositories
- No-code data pipeline automation
- 24-hour support and error recovery

But security is the most crucial aspect of any ETL solution. If a vendor can't offer a full suite of security options, then it's worth shopping around for someone you trust.

Key data security features

As discussed in the previous section, there are certain things that you should verify about each vendor, such as SOC compliance, physical security and reputation.

It's also a good idea to look at the data security features they offer to users. The most important ones are:

- **Secure login**

  Your team will access the Cloud ETL service through a web interface. This interface should offer a secure connection with outstanding authentication features, including 2FA and suspicious activity detection. Do they offer Single Sign On (SSO)?

- **SSH/Reverse SSH tunnel**

  The best ETL vendors will allow you to connect without compromising your security. This usually involves an SSH tunnel or reverse SSH if you can't provide port access. Integrate.io supports both SSH and reverse SSH.

- **Non-persistent data**

  ETL should transport your data from A to B with no records in-between. This means no copies, no archives, no logs – nothing that might inadvertently cause a risk of a data breach. Look for a service like Integrate.io that guarantees the non-persistence of all data passing through the pipeline.

- **Data encrypted in transit, and at rest**

  Within the ETL process itself, data is sometimes at rest or moving between locations. The vendor should be able to guarantee robust encryption for in transit and at rest throughout the ETL process.

- **Regular penetration testing**

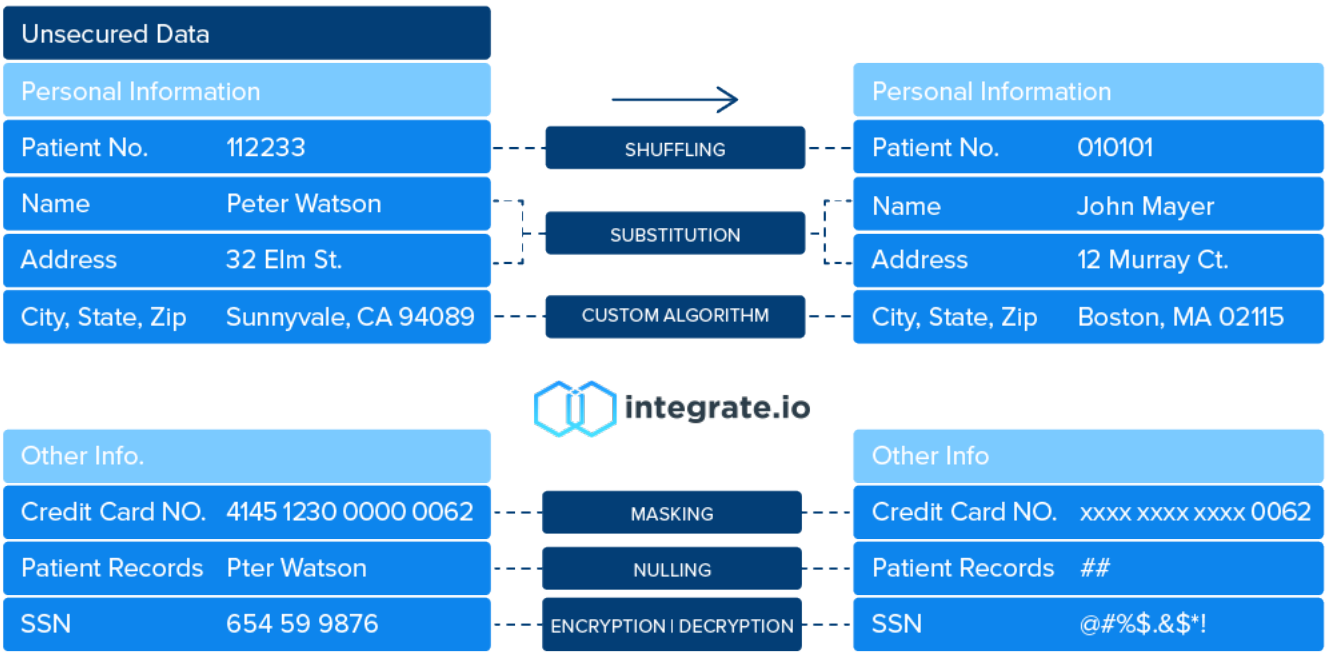
  As per SOC 2 requirements, penetration testing occurs regularly. You might request the PenTest reports when signing up, but remember to keep checking them each year while you're signed up for the service.

Any cloud ETL service should offer these as a minimum.

Security through data transformation

ETL can also improve your overall level of data security by offering transformation functions that protect sensitive data.

- **Field-level encryption:** Field level encryption means that data is always encrypted when it leaves your network. Decryption is impossible without the key, which you hold on your side. Should anyone intercept or access data while it's outside of your network, they won't be able to decrypt it. Integrate.io offers field-level encryption using Amazon's Key Management Service (KMS), and you can use this to encrypt data at any stage in the ETL process.
- **Hashing:** Hashing is a one-way cryptographic function that replaces sensitive data with a meaningless value. For instance, you can configure your ETL to replace social security numbers with a set of random characters.
- **Masking:** Masking is commonly used in testing and analytics scenarios, where your team might need large volumes of representative data, but they don't need genuine personal information. An ETL masking layer will produce an arbitrary value that meets requirements but doesn't expose personal information. For example, the ETL platform could replace Social Security numbers with a random 9-digit number.
- **Obfuscation:** Obfuscation is a way of hiding data values that is often reversible. For instance, the ETL may replace certain values with codes from a lookup table. That lookup table later makes it possible to restore the original values.

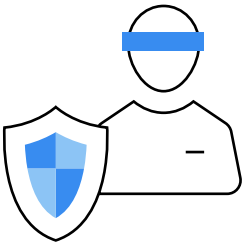


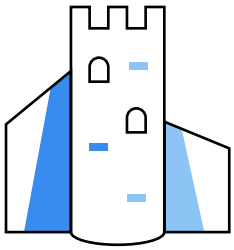
ETL can help to minimize risk to data in transit by hiding or removing sensitive information. This is the ultimate data security strategy – if someone does intercept or access the data, they won't find anything of value.

08

6 SECURITY QUESTIONS TO ASK YOUR ETL VENDOR

The right ETL vendor can have a massive impact on your overall level of data security. As with any professional partnership, it's essential to get off on the right foot. You can do this by looking at the product features and asking whether they truly align with your needs. More importantly, try to have a conversation with your vendor. Talk to them and see if they understand your needs. Here are a few questions to ask when you approach a vendor.





1. How can your platform help protect our PII, PHI, and other sensitive data?

There's no one-size-fits-all approach to data security, so your vendor shouldn't offer a one-size-fits-all answer to this question. Instead, they'll talk to you about your particular needs and explore issues like:

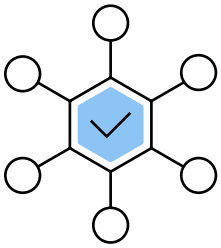
- What kind of sensitive data do you collect?
- How do you store and process this data?
- What territories do you operate in?
- Who is using the data?
- What kind of production systems and storage solutions are you using?
- What are your analytics objectives?

This will help them understand your needs and also to identify potential risks in your data strategy. The vendor should then be able to suggest ways that their ETL solution will be able to help.

2. What examples can you share of how you have helped other clients with their data security?

Many ETL vendors have worked on big projects for major organizations. They'll have experience with complex data infrastructures, and they'll know how their solution can address real-world problems.

Ask your vendor for case studies and testimonials to show that they have this kind of background. This will allow you to gauge their reputation and see if you're working with someone you can trust. Case studies will also let you know if they have dealt with organizations like yours in the past.



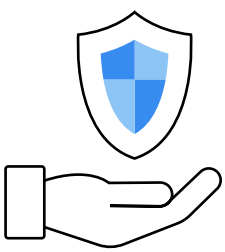
3. What features does your platform have to maintain compliance with regulations such as GDPR, CCPA, HIPAA?

Any reputable vendor will already be compliant with all major regulations. For example, Integrate.io ETL meets the requirements of GDPR, transforming data in the EU, and offers an updated Data Processing Addendum (DPA) to support customers' GDPR compliance needs.

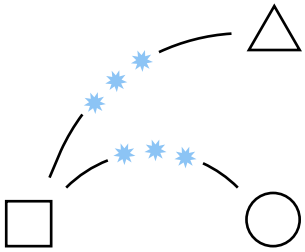
Your vendor should understand how regulations might impact you and offer advice on how to stay compliant. Remember – you are the responsible party if your third-party ETL service causes a compliance breach. Protect yourself by choosing a partner that understands the law.

4. How can your data security team assist with our data security strategy and implementation?

If your ETL vendor has a security-first mindset, they'll be able to offer advice and suggestions about keeping your data safe. They may offer some resources on building an effective data strategy, as well as guidelines on meeting standards such as SOC 2.



The simplest way that ETL vendors can help is by offering a secure one-to-one data pipeline between systems. This is much more secure than the many-to-one architecture of some infrastructures.



5. How do you remove/encrypt sensitive data in Europe for GDPR before moving data to the U.S. or elsewhere for centralized analysis?

Moving data across national borders is increasingly tricky in terms of compliance. Unfortunately, most organizations need to move data internationally. Even if you don't have an office abroad, you might use an accounting, analytics, or storage service based in another country. Sending data to them could put you in breach.

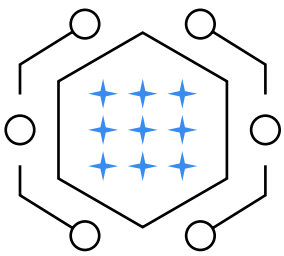
ETL makes things much easier by offering tools such as data obfuscation and field-level encryption. These transformations (performed in an EU data center) can make data compliant before transit. You can then allow your data pipeline to run as normal without worrying about breaches.

6. Does your platform support field-level encryption for sensitive data fields?

Field-level encryption is the most secure way to protect personal information. Encryption happens before data leaves your network, and there's no way to decrypt it without the relevant key. It's a failsafe system – if hackers manage to access your data, they won't be able to interpret it.

It's important to ask whether your vendor offers field-level encryption. Also, ask them:

- Which encryption service do you use?
- How does your ETL handle encryption?
- How do I encrypt and decrypt data in the pipeline?



With Integrate.io, you can encrypt and decrypt from the expression editor, using the Encrypt() and Decrypt() commands.

09

CONCLUSION

In a global, digital world, the most important currency for any business is trust.

Trust is something you earn over many years by diligently safeguarding your customer's personal data. They trust you with a lot of essential information – their address, their payment details, their preferences, even their biometric data.

But trust is something you can lose in an instant. All it takes is one lapse in data security, and nobody will ever trust you again. They'll switch to a rival that takes data security seriously, someone that always protects customer confidentiality.

Data security isn't easy. Cyberthreats are constantly evolving, and employees struggle to keep up with ever-changing protocols. The new normal of remote work has added another layer of risk to an already challenging world.

This is why it's more important than ever to get data security right. It starts with strategy, with people and with education. But the most important part is getting the infrastructure right. A good ETL vendor can help you build a secure data pipeline that keeps sensitive information out of the wrong hands. Your customers will rest easy knowing their personal data is safe.