White Paper

Seizing the cloud opportunity — safely and securely

Data protection in the Equifax cloud native architecture – built on the Google Cloud Platform (GCP)

August 2020

This paper provides a high-level overview of the Equifax cloud native transformation on GCP, and details how data protection remains prioritized and delivered across the architecture.

The Equifax cloud native transformation

The Equifax cloud native transformation is the largest investment in the company's history and necessary to deliver on the company's long-term strategy. Equifax expects this migration to transform its business, accelerate growth and allow the company to provide unique data, products and services that can be only delivered in a cloud environment.

The company primarily leverages Google Cloud Platform (GCP) for its transformation. The overall objectives of the transformation include:

- Securely store, process and find value in a continuous volume of structured and unstructured data
- Continually improve and personalize experiences
- Rapidly design, deploy and maintain products, services, and data
- Grow business with smarter decision making while constantly adjusting to a changing world of data privacy requirements
- Provide "Always On" availability to our end users
- Leverage scalable services for dynamic workloads
- Provide low latency and lightweight API calls to ensure rapid response times for complex decision responses

Equifax also understands the particular importance of built-in security controls for sensitive data and systems. This paper focuses on a critical aspect of those controls, specifically, those around data protection.

Equifax understands the particular importance of built-in security controls for sensitive data and systems

But before turning to data protection, it is important to keep in mind two key aspects of the Equifax cloud native transformation: reducing complexity and increasing speed of innovation.

Reducing complexity by becoming cloud native

Equifax believes that in most cases it is quicker and easier to simply rebuild legacy IT applications in a cloud native form than to slowly migrate them through a hybrid journey over time.

Traditional business applications are complicated and expensive, especially at scale, because the amount and variety of hardware and software required to run them are daunting. It can require a large team of experts to install, configure, test, run, secure and update them. Multiply this effort across dozens or hundreds of apps and it's easy to see why the biggest companies with the best IT departments are not getting the apps they need. Small- and mid-sized businesses do not stand a chance. In addition, many companies are drowning in technical debt and limitations.

When faced with this reality, the traditional approach is a long journey that leverages a hybrid computing model. This approach does not solve the problem quickly, and in many cases exacerbates the problem. IT teams are left with a stack of incomplete hybrid implementations and must operate both the traditional business applications and the hybrid ones. This often accomplishes nothing but a rise in complexity and operating costs.

Given the challenges of traditional business applications and the realities of maintaining those applications or a hybrid implementation, a new wave of technical thinkers are cracking open the problem and forcing it to be solved. These developers believe that the cloud native business applications they are writing today are creating value over the long term and will ultimately achieve substantial benefits at a reduced cost over the traditional and hybrid applications today.

The Equifax difference: Increasing innovation by breaking down the barriers of data flow

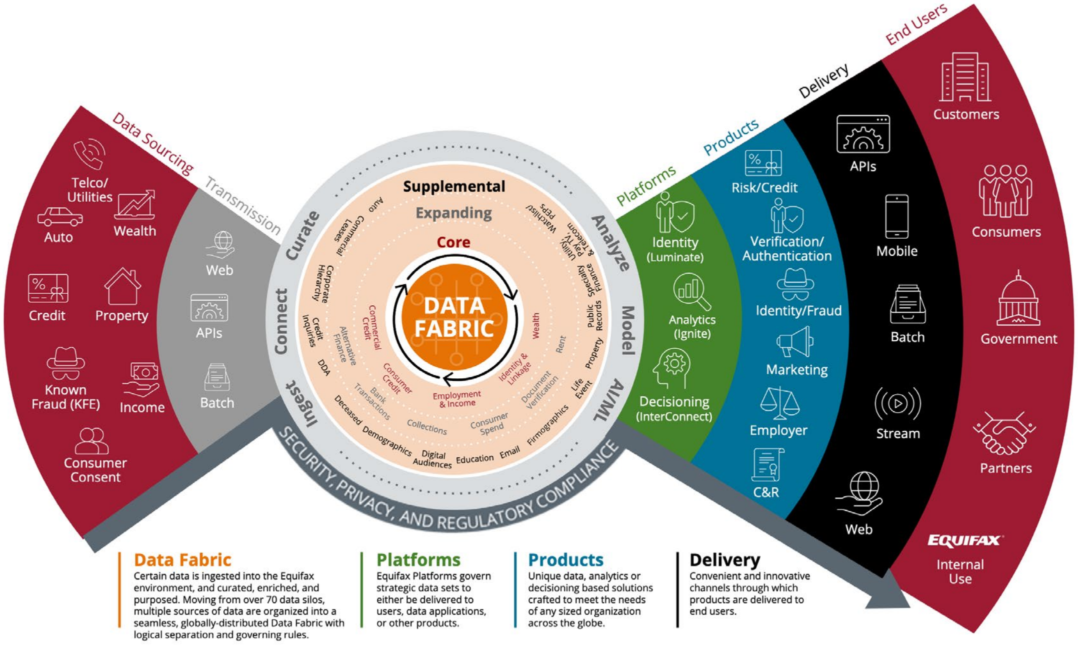
The Equifax transformation will result in a set of cloud native tools that enable developers to rapidly innovate and data scientists to spend less time sifting and sorting and more time investigating and deciding.

It is this difference – the Equifax difference – that is simple yet powerful: smarter insights for smarter action.

The Equifax transformation will result in a set of cloud native tools that enable developers to rapidly innovate.

To achieve this difference, Equifax has broken down the silos that existed among various engineering teams and is rebuilding its data flow into fundamental building blocks, aligning engineering teams to those blocks, and developing the right platforms and services to best provide value to our customers.

Below: The Equifax structure has been simplified to break down the barriers of data flow.



Data Fabric
Certain data is ingested into the Equifax environment, and curated, enriched, and purposed. Moving from over 70 data silos, multiple sources of data are organized into a seamless, globally-distributed Data Fabric with logical separation and governing rules.

Platforms
Equifax Platforms govern strategic data sets to either be delivered to users, data applications, or other products.

Products
Unique data, analytics or decisioning based solutions crafted to meet the needs of any sized organization across the globe.

Delivery
Convenient and innovative channels through which products are delivered to end users.

This new structure helps the Equifax network make better decisions and gives engineers and data scientists the ability to easily and powerfully mix existing applications with available datasets. It does this through highly scalable, available and dynamic platforms. The operational efficiencies gained by shifting to the cloud should also increase customer service levels.

Data protection: The Equifax approach

There are many methods to protect sensitive data including:

- Sound data governance policies
- Network and infrastructure changes, which include network firewalls, Intrusion Prevention Systems (IPS), and semantic layer row- and column-level security (RLS/CLS)
- Identity and access management policies and procedures, including Identity Management (IDM), Role-Based Access Controls (RBAC) and activity monitoring
- Data-focused methods, which include encryption and hashing/HMAC

This paper focuses on the Equifax general approach to data protection as it migrates to the cloud.

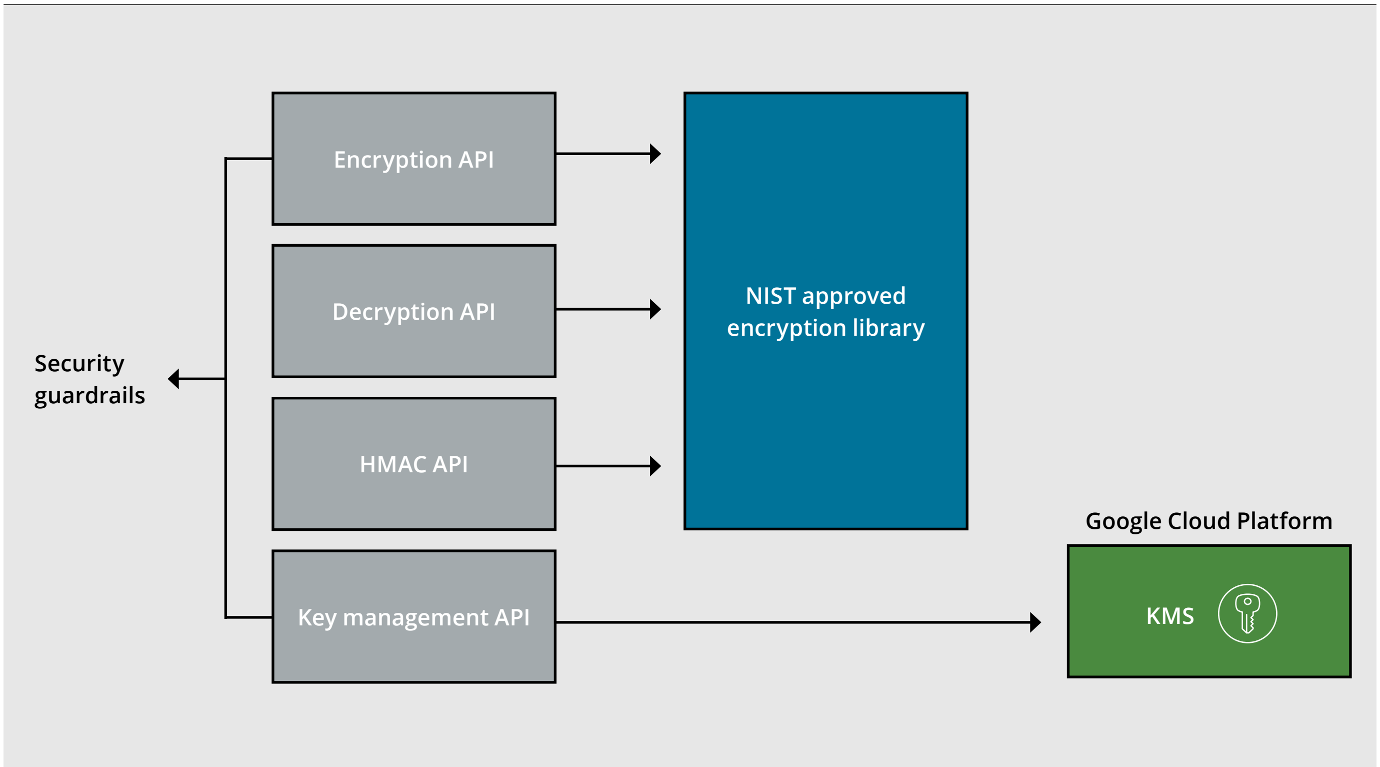
Data protection strategy

While cloud vendors offer highly secure environments, their customers are responsible for the data stored within them. (This is commonly referred to as the Shared Responsibility Model.) The Equifax data-first approach goes above the standard at-rest encryption policies available by the majority of public cloud providers. Additionally, Equifax has to ensure that its data protection toolkit for its cloud implementations is as flexible as possible.

A key item in the Equifax toolkit is the requirement to implement field/record level encryption for the most sensitive data. This is accomplished through the use of NIST-approved cryptographic libraries. To enforce good cryptographic hygiene, Equifax has developed a proprietary library consisting of APIs used to add guardrails around the consumption of the base cryptographic module.

> The Equifax data-first approach goes above the standard at-rest encryption policies available by the majority of public cloud providers.

Equifax data protection library



The Equifax data protection library addresses the following cryptography items:

- Implementation of industry best practices
- Best-in-class key management
- Enterprise-wide unification of encryption algorithms/modes
- Secure key/data sharing mechanisms
- Data rehydration/re-encryption
- Secure search and match solutions via HMAC

For specific use cases, Equifax has the ability to allow customers to maintain full control of their data usage rights for data encrypted at the field or record level by integrating Customer Supplied Encryption Keys (CSEK). This implementation of CSEK allows for split key ownership in which the entire key is never owned by one party. In the event that the customer no longer consents to share their data with Equifax, they have the ability to restrict access to their key shares, thus crypto-deleting their data.

Data categorization

Equifax maintains a robust retention and removal program, and ensures that data is removed when no longer needed under our retention standards. Any data that is required to be retained is categorized into:

- Internal search key data
- Input specific/reversible data

For internal search key data that is solely used for matching purposes, a special type of hash that incorporates a data encryption key may be used (HMAC). The outcome of the process is deterministic, but results in a one-way cryptographic hash that is virtually useless without the original input.

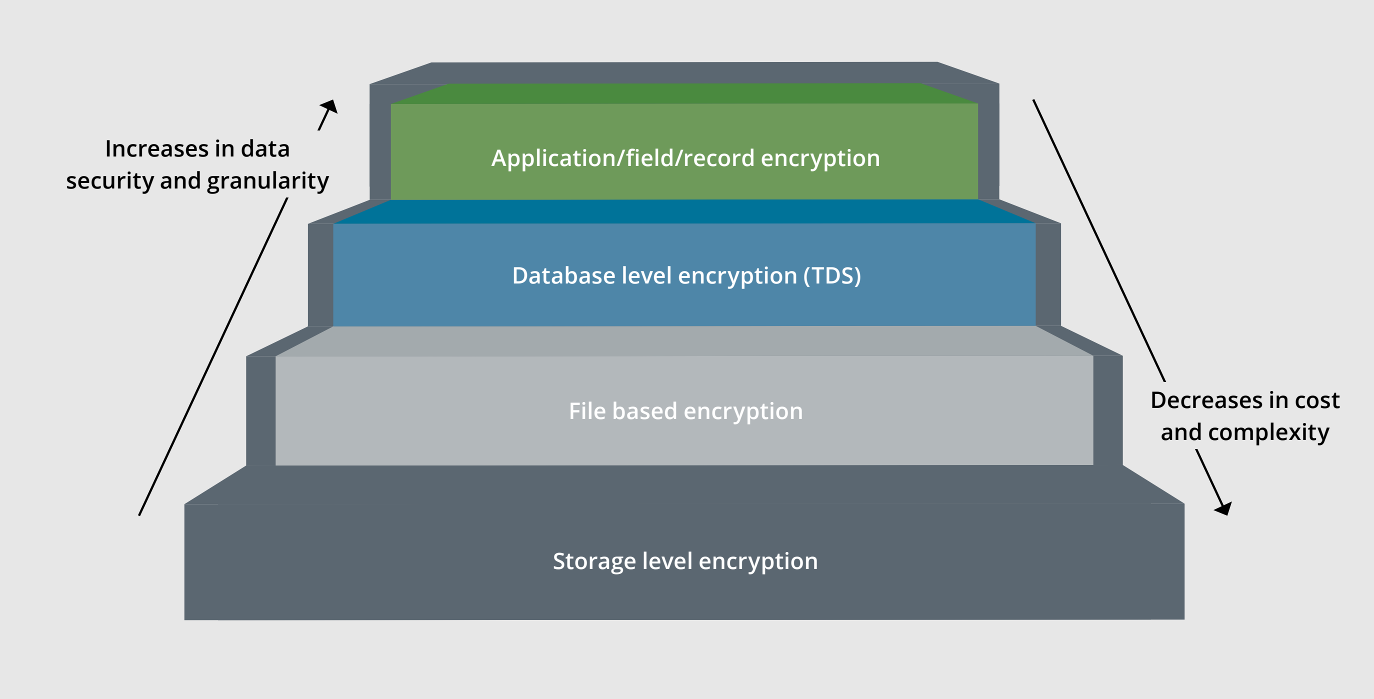
As we continue to migrate data into the cloud, a reversible form of encryption is required for all other sensitive data deemed necessary for business and analytical operations. The encryption process utilizes mathematical functions and cryptographic keys to transform data into binary ciphertext which differs greatly from the original input. In order to revert the ciphertext back to its original form, the user must call the data protection library and supply the corresponding data encryption key (DEK) used during the encryption process.

Data at rest

Whether data is stored in a database, file system, private data center or a public cloud storage device, there is always an inherent risk of exposure. There are varying levels of encryption that can be applied to mitigate these risks. Typically the more secure the implementation method, the more complex the effort becomes.

Equifax maintains a robust retention and removal program, and ensures that data is removed when no longer needed.

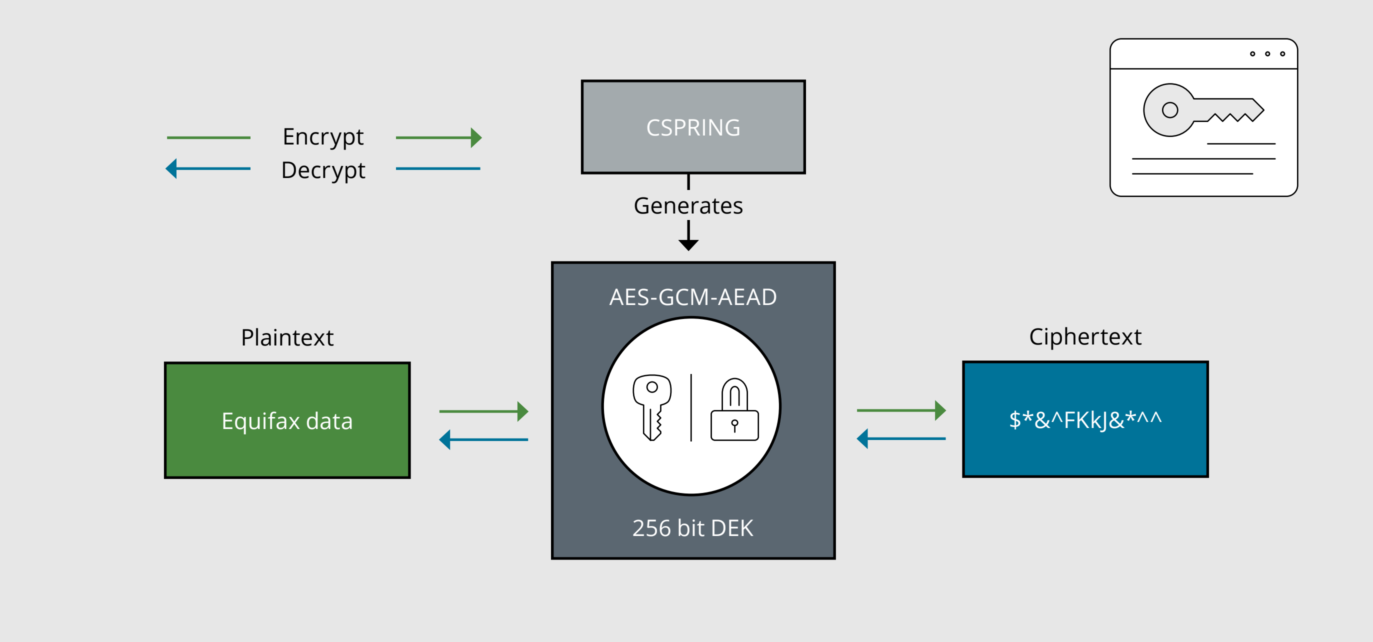
Varying levels of encryption


Increases in data security and granularity
Application/field/record encryption
Database level encryption (TDS)
File based encryption
Storage level encryption
Decreases in cost and complexity

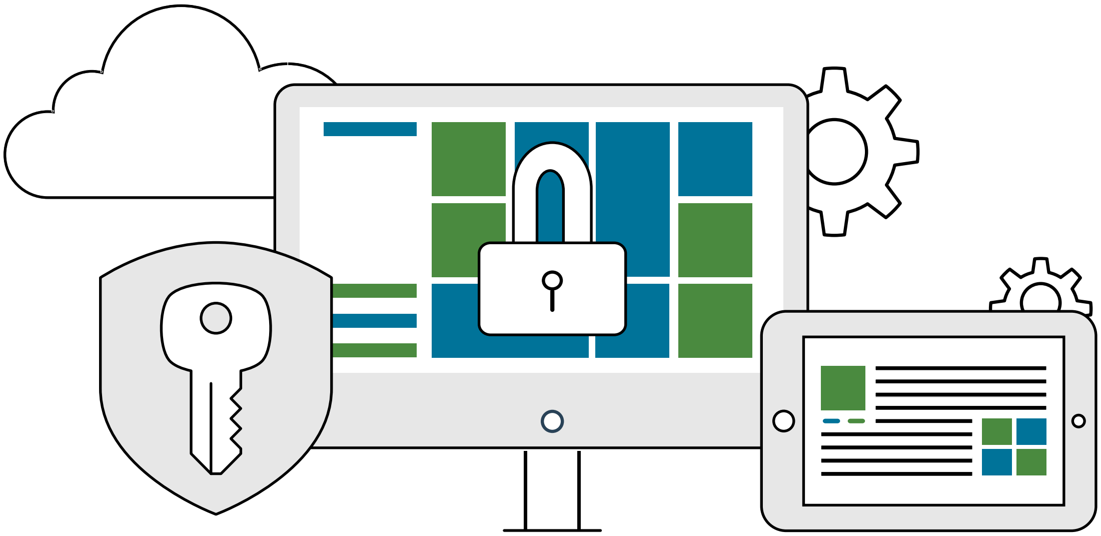
Equifax uses all methods described in this illustration, in combination with other security controls. Different use cases require a different balance between risk and complexity. Our data protection library is designed to allow us to achieve the top level of the data protection illustration depicted above for our most sensitive data in our cloud environment.

The algorithm used by our data protection library is Advanced Encryption Standard (AES) utilizing a 256 bit encryption key in Galois Counter Mode (GCM) with optionally authenticated encryption in the form of Authenticated Encryption Additional Data (AEAD) resulting in a non-deterministic output. Traditionally encryption has been used to encrypt data at the file level; however, with advancements in cloud data storage technologies, the capability to accommodate ciphertext of varying lengths and outputs now exists.

Advanced Encryption Standard (AES)



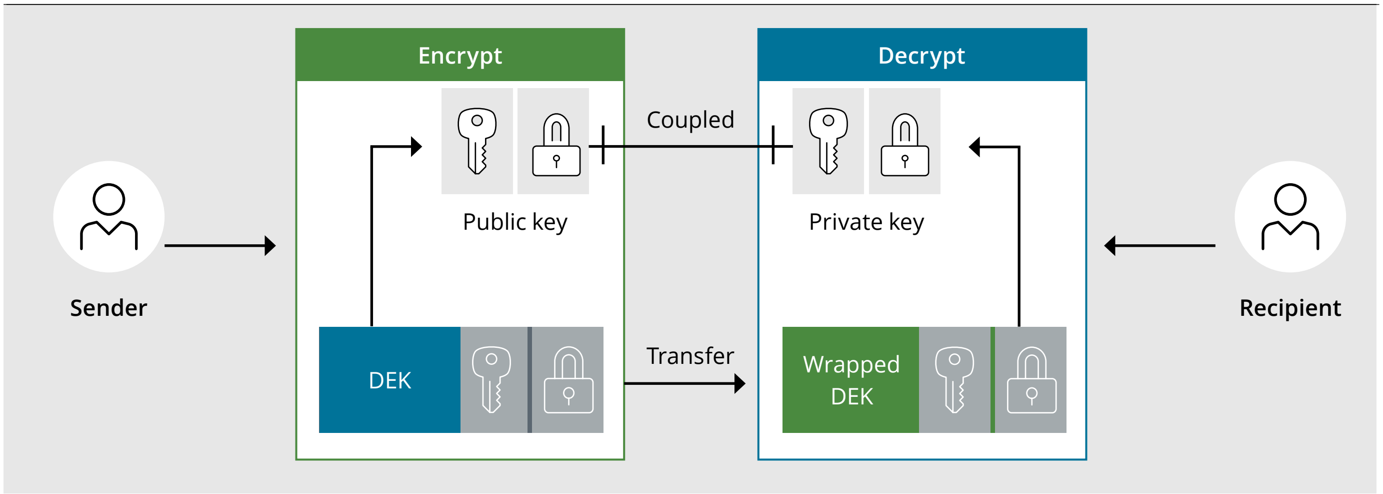
Due to the non-deterministic nature of the Equifax encryption implementation, a combination of encryption and HMACs are used to alleviate the need to decrypt entire datasets to perform search and match operations.



Data in transit

Equifax relies on network encryption protocols (e.g., TLS, SFTP) to secure the channel that is being used to transport data across a network. In addition to securing the transmission channel, the data inside the channel is also secured with a mixed use of asymmetric and symmetric encryption. Due to the fact that asymmetric encryption is severely less performant than symmetric encryption, it is industry standard to use asymmetric keys to encrypt the symmetric data encryption key (DEK). This is used to encrypt the payload/message. By using this approach, the sensitive data is not exposed – even if the secure connection is in any way terminated or compromised. Moreover, by standardizing the data in transit strategy, both the sender and recipient are able to rely on the same enterprise data protection library for safe and reliable data transmission.

Data in transit



Data usage

Data access is granted on a least privileged basis. That is, only authorized users will be able to view decrypted sensitive data that is related to their job functions. Moreover, the application layer encryption policies achieved through the data protection strategy prohibit administrator level users from viewing data not required to complete their job functions. This is not possible with the native public cloud offerings alone.

Next, separation of duties is just as important as protecting the data itself. By removing the key creation process from the individuals and packaging that into the data protection library, three things are accomplished:

1. Cryptographic keys are created with consistent entropy
2. Sensitive key material is not exposed to the end user
3. Wrapping Keys/Key Encryption Keys (KEKs) are controlled by security officers instead of application teams

Access to the wrapped keys and data is never controlled by an individual for security purposes. In order to gain access to plaintext data, a service account must be able to retrieve the data encryption key, decrypt the wrapped DEK using the KMS, retrieve the encrypted data and make calls to the data protection library to perform the operation. Federated logins are used to control the access to service accounts in which periodic entitlement reviews are conducted to maintain access validity. It is also important to note that the data protection library ensures that while handling DEKs, the key material is only held in volatile memory for security purposes; further, the APIs in the library do not allow the user to return the key material in its unwrapped form.

By standardizing the data in transit strategy, both the sender and recipient are able to rely on the same enterprise data protection library for safe and reliable data transmission.

Implementation

The Equifax data protection strategy is designed to be compatible with massively parallel data processing frameworks allowing for millions of operations per second while maintaining a high degree of modularity. By breaking the cryptographic components down into smaller pieces, it is possible to replicate the same algorithms across a wide array of services. For example, one implementation of the data protection library is used during the ingestion process while the same cryptographic equivalent is used for securing online credit information. This can also be extended to collaborative efforts with our partners to ensure that the same levels of encryption are used in their native cloud offerings (e.g., GCP BiqQuery's native user-defined functions).

Due to the performance implications when dealing with large scale encryption operations, it is important to minimize the amount of times data is reidentified. In a typical data flow, data is accessed upwards of 10 times before it reaches the intended data store. Such actions include address and name standardization, keying and linking, aggregation, filtering and data purposing. Further, by protecting data at a fine-grained level, the user is able to safely and more effectively access only the pieces they need to complete their part of the data processing pipeline. This is where field level encryption has the advantage over other more coarse-grained encryption strategies such as file level encryption. By protecting data at the field level, the person accessing the data is limited to only the attributes (e.g., name and address) and nothing more. This keeps the rest of the data protected and also cuts down on the amount of processing required to complete the task at hand.

By protecting data at the field level, the person accessing the data is limited to only the attributes (e.g., name and address) and nothing more.

Interoperability

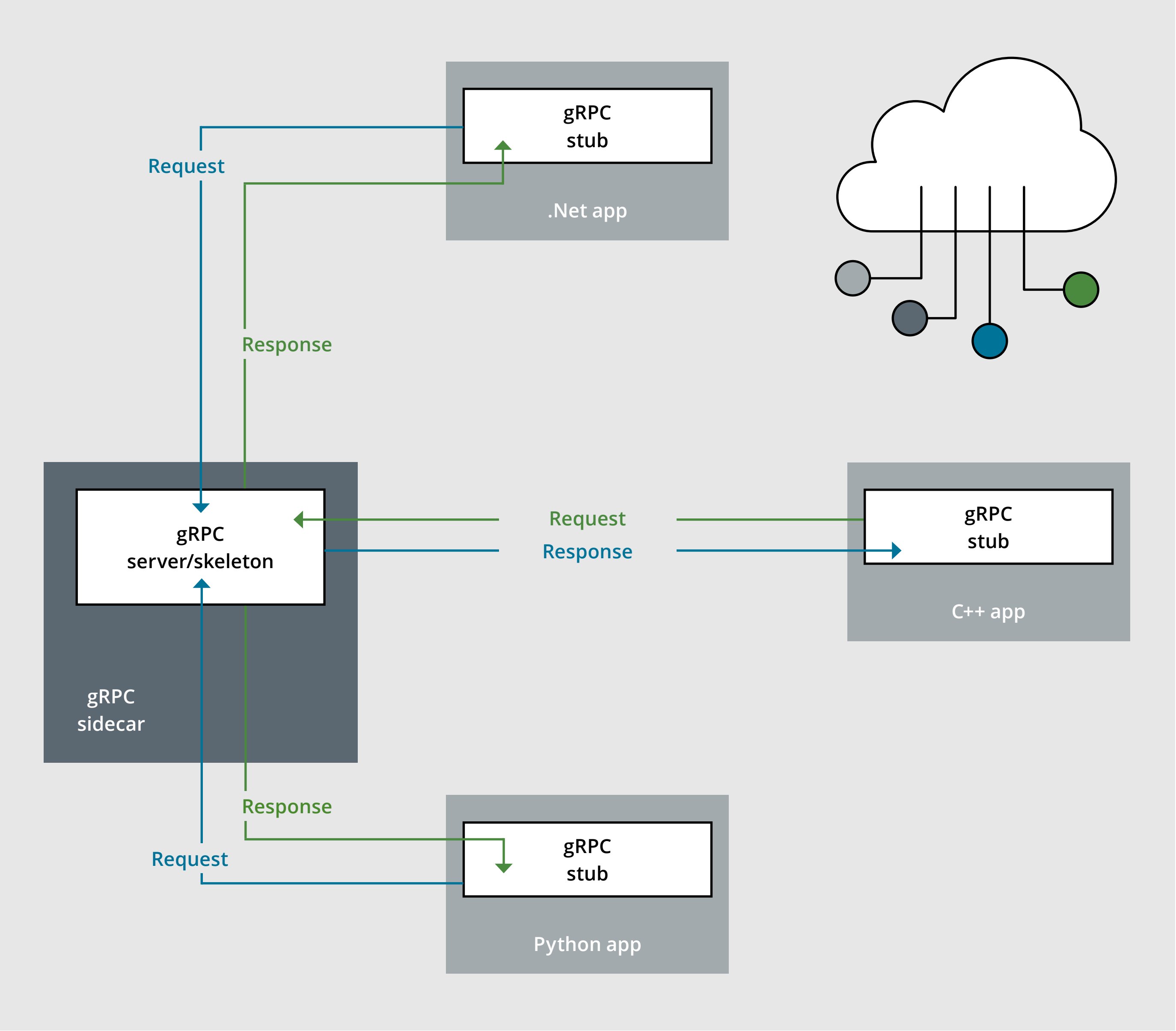
Another important design aspect to consider when designing large scale encryption systems is the ability for systems to pass protected data without having to decrypt in the middle. When system A sends data to system B, there are three main goals to consider:

1. Secure data transfer
2. Minimal processing
3. Multi-language support

To accomplish the secure data transfer, asymmetric encryption is involved which was discussed in the previous section. The harder piece to solve for is the minimal processing which if done correctly results in true interoperability among your applications. Agreeing on and incorporating an enterprise standard is the easiest way to achieve interoperability. Once the stakehokders agree on an algorithm and mode, a standardized library is built for teams to consume. This way when system B receives the protected data from system A, they are able to work with the data without tons of post processing or cryptographic overhead.

In order to accommodate the plethora of coding languages used across the enterprise, a gRPC sidecar implementation of the original data protection library was introduced. The sidecar is designed to run in a containerized environment in which the user makes calls locally using one of the 14 supported languages. All of the services are predefined and the interpretation layer is handled by gRPC. This increases the flexibility of the enterprise data protection strategy as application teams are not forced to rewrite applications or waste time developing their own language specific cryptographic libraries.

Data in transit



Future state

As Equifax continues our cloud transformation initiative, our goal is to develop data-first solutions in which Equifax and public cloud services are able to integrate seamlessly while maintaining the highest degree of data protection and integrity. In order to fully accomplish this, Equifax is continuing to partner with public cloud providers to stay up to date with NIST cryptography standards and how Equifax can help each other provide seamless integrations with as little rework as possible.

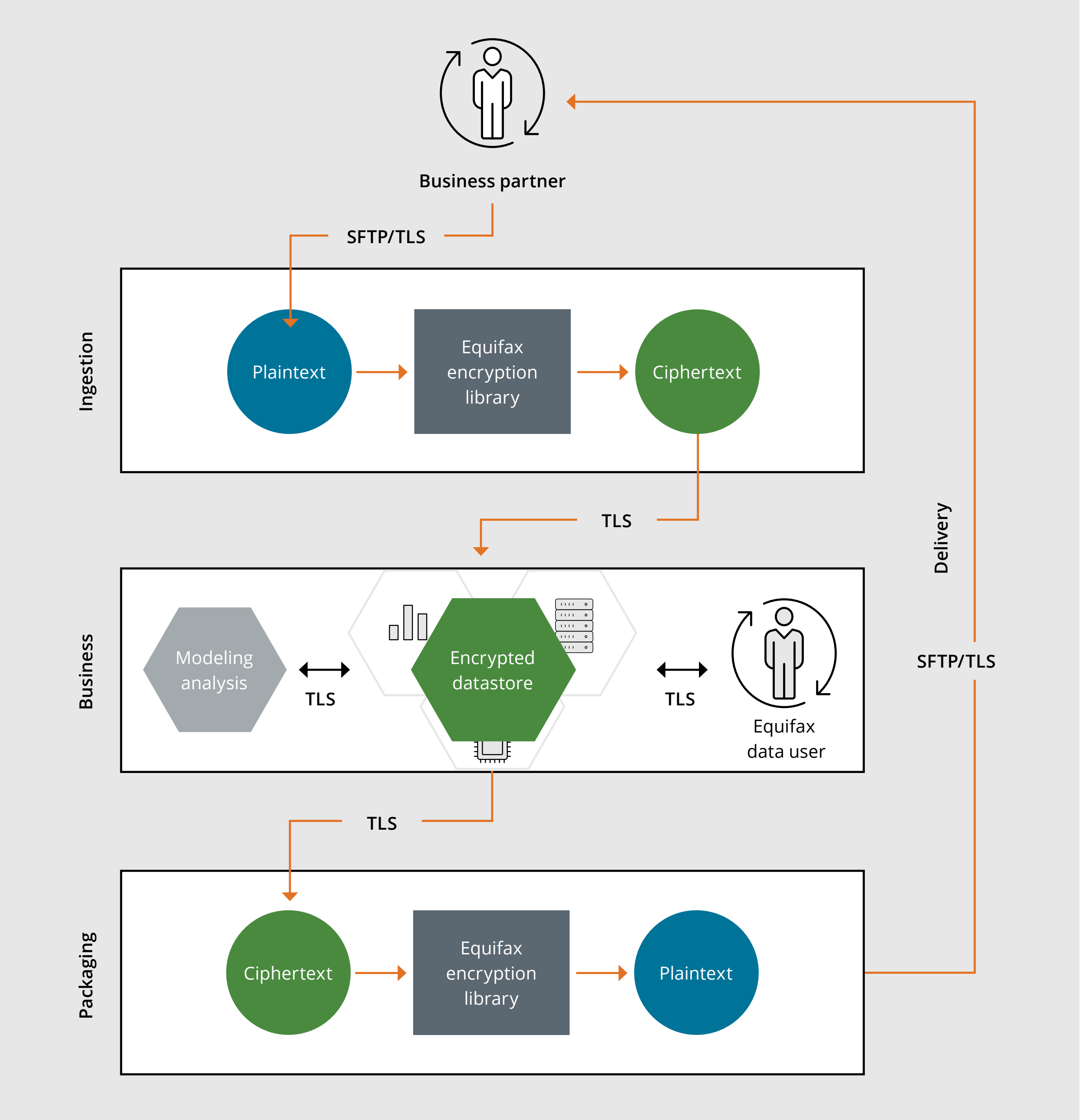
An ideal flow would look like:

- Data is encrypted during the ingestion phase
- Data flows to each system without any further cryptographic processing
- Data users are natively able to perform business functions needed from any public cloud service offering without custom cryptographic user defined functions
- Data is decrypted and made available for packaging prior to delivery

Following this process, data remains protected throughout the lifecycle processing within Equifax boundaries.

Equifax is continuing to partner with public cloud providers to stay up to date with NIST cryptography standards.

Lifecycle processing within Equifax



Conclusion

As Equifax transforms to a cloud native environment, it is committed to protecting customer data, not just because Equifax is mandated by regulatory bodies, but because it is the right thing to do for the global economy. By putting data first, Equifax is committed to safeguarding customers' data.

equifax.com/about-equifax/security-technology-transformation