Security whitepaper

01/13/2020

Table of content

Introduction 4

Infrastructure security 4

Cloud provider and location 4

Technology stack 4

Storage 4

Security-first culture 5

Source code development 5

Vulnerability prevention 5

Secure access 5

Operational security 6

Operational monitoring and failover 6

Incident management and notifications 6

Isolation 6

Application security 6

No direct access 6

API 7

Integrations 7

User access 7

User security 7

Authentication 7

Credit cards 7

Data protection 8

Securing data 8

Encryption in transit 8

Encryption at rest 8

Use data lifecycle 8

Data removal 9

Secure environment variables 9

Deployment keys 9

Access rights system 9

Vendor management 10

Additional resources 10

This whitepaper provides detailed information about security measures, standards, and best practices applied at Apify to protect its customers' data. The whitepaper further discloses the main cloud providers, data center location, and technologies used.

Introduction

Apify provides a cloud computing platform to download, extract, and process data from the web or automate workflows on the Web. Users ranging from small startups to Fortune 500 companies trust Apify to provide a secure and reliable platform without exposing their data to risk. At Apify, security is the top priority of our daily work. Security best practices are reflected in our development, deployment, monitoring, and project management processes

Infrastructure security

Cloud provider and location

The Apify platform's core services run on Amazon Web Services (AWS), a certified enterprise-grade cloud infrastructure provided by Amazon.com, Inc. The core service servers and storage systems are located in Northern Virginia, United States (the **us-east-1** region). The servers and storage systems are protected using industry-standard best practices, such as IP filtering, limited remote access, and enforcement of encryption.

Apify also utilizes a content delivery network (CDN) and proxy servers that are physically located in a number of countries outside of the United States, but no user data are stored on these systems. By default, all communication with these systems is encrypted.

Technology stack

Apify platform is built on a highly scalable technology stack developed using Node.js, Docker, and Linux (Ubuntu). It runs on top of various AWS services such as EC2, S3, DynamoDB, SQS, ECS, Lambda, and CloudFront.

Storage

The primary database containing user data is a MongoDB cluster, which is hosted by the MongoDB Atlas service provided by MongoDB, Inc (https://www.mongodb.com/cloud/atlas). The cluster also runs in the AWS **us-east-1** region, the data is encrypted at rest, and all communication between the database and other systems is encrypted. Operational data in Apify Storages such as Request queues, Datasets, and Key-value stores are stored using AWS services such as DynamoDB, ElastiCache, and S3.

Security-first culture

Source code development

The Apify platform's source code is developed in-house by the Apify team using industry-standard practices. Each code change is reviewed by at least two other developers for both functional and security requirements before being merged into the main branch. All code changes are rigorously tested using unit and integration tests, and deployed using a Continuous Integration (CI) pipeline.

The Apify team follows best practices to develop secure software and maintain a secure infrastructure. We use higher-level programming languages such as JavaScript that reduce possibilities of lower-level security vulnerabilities like buffer overflows. Further, we do not construct any SQL queries, preventing the chance of SQL injection attacks.

Vulnerability prevention

The Apify team monitors all relevant security bulletins (e.g. for Amazon AWS, Linux, Ubuntu, and Node.js) and uses automated monitoring of vulnerabilities in external packages and libraries that are integrated into the codebase. For example, the "npm audit" tool is called after each merge to our Node.js codebase to scan the project for vulnerabilities.

EC2 instance images, Docker images, tools, all packages, and all additional software are regularly updated for both functional and security updates.

Secure access

All Apify employees and contractors are given the least privileged access to both internal and external systems required to productively perform their job. Apify employees always use individual rather than shared accounts and enforce two-factor authentication whenever possible (for security-critical systems, always).

Operational security

Operational monitoring and failover

All infrastructure resources such as servers and databases have replicas for redundancy and are automatically monitored for failures. In the case of an outage, the workload is automatically switched to the replica. Logs from all systems are collected in a centralized store provided by LogDNA, Inc. and automatically monitored for errors.

All user-facing APIs and features are automatically and periodically tested with a battery of integration tests.

Incident management and notifications

Systematic failures or outages of any component of the Apify platform are automatically reflected on the Apify platform status page (https://status.apify.com), enabling all subscribed users to be immediately notified about the incident. The Apify team updates the status of the incidents to keep users informed about the root cause of the failures, the impact, resolution, and steps taken to avoid such failures in the future.

Isolation

Tasks executed by each user run in isolated environments in terms of computational, storage, and network resources, and cannot directly interfere with each other and access each other's data.

Application security

No direct access

No user of the Apify platform is provided with direct access to the database or any of the underlying servers or storage systems used by the platform. Instead, all external access to user data is facilitated using Apify's RESTful HTTP-based API, which ensures authentication, granular access, and encrypted communication.

API

User access to the Apify API is authenticated using access tokens, which are random strings generated using a cryptographically-secure random number generator and are long enough so they cannot be guessed by any other party. The communication between clients and API is secured using HTTPS (SSL/TLS) encryption.

Integrations

All external integrations of the Apify platform (e.g. for Zapier or Integromat) are implemented using the secure Apify API and are thus secured with the same measures as the API itself.

User access

Users are authenticated using the methods described in the “User security” section and all communication leverages HTTPS/TLS encryption

User security

Authentication

Users can create an Apify account either by providing a password or by linking their Google or GitHub account. Passwords are never stored in a plain text or any form that enables reconstruction of the password. Instead, Apify only stores a so-called hash of the password, generated using the industry-standard Bcrypt algorithm. The passwords need to be at least 9 characters long and contain both numbers and letters.

Credit cards

Apify never has access to credit card numbers of users and only uses PCI-compliant vendors (PayPal, Inc.) to process credit card data. The credit card numbers entered by our users during subscription are passed directly to the PCI-compliant providers via an IFRAME, and Apify never has access to them.

Data protection

Apify is deeply committed to providing its users and customers with maximum security and privacy and is committed to complying with the European Union's General Data Protection Regulation (GDPR). For more information, visit https://apify.com/gdpr.

Securing data

Encryption in transit

All user data that is transferred via public networks is encrypted using industry-standard strong encryption SSL/TLS protocols. The encryption is applied to all integrations with external systems, web applications and APIs.

Encryption at rest

The primary MongoDB database that stores user data is encrypted at rest including the database snapshots. In addition, all user data is stored securely with access limited only to systems that need access to them. Users can access the data only via API servers that authenticate them.

Further, user secrets such as secure environment variables and deployment keys are stored encrypted using a public key. They are only decrypted using a private key when the secrets are needed by the system.

User data lifecycle

User account settings, actors, actor tasks, and named storages are persisted until the user removes their account. Operational data such as actor (task) runs, unnamed storages, and logs are removed after the specific user's data retention period, which is based on their subscription plan. Note that it can take up to one month after the resource was deleted by the user or at the end of the data retention period, before the underlying data is completely erased from all the storage systems and caches.

Data removal

Any user can remove their account at any time along with all the data associated with the account. This functionality is available under the settings tab of the user Account page at https://my.apify.com/account. Note that it may take up to one month for all the data to be completely removed. Apify will retain only the minimum amount of information required for accounting and legal purposes.

Secure environment variables

User secrets that are needed to implement custom solutions using Apify Actors (https://apify.com/actors), e.g. external API tokens or passwords for website login, can be saved as secure environment variables. These variables are stored encrypted using a public key in Apify databases and only the systems that require them are provided with a private key to decrypt the values. These secrets can neither be retrieved by API nor via the Apify app user interface.

Deployment keys

Deployment keys are used to deploy the code from a user-owned Git repository. The same measures as for secure environment variables apply also for deployment keys. Keys are stored encrypted using a public key and cannot be retrieved by an API or via the Apify user interface.

Access rights system

By default, users can only access their own data. Users can share access to their own content with other users using a granular access rights system (https://docs.apify.com/access-rights).

Vendor management

Apify uses a number of external services from various companies in order to provide its own service to users. These services are used for a large number of purposes, ranging from tools for communication with customers provided by Intercom, Inc. to web user interface error monitoring tools provided by Sentry, Inc. Apify only works with reputable external services and companies that meet our high selection criteria and use the best industry standards for security and quality of serviceApify follows the status pages and incident reports of these services to evaluate how their potential incidents could affect the Apify platform and its users.

Additional resources

- Amazon AWS security - https://aws.amazon.com/security/
- MongoDB Atlas security standards - https://www.mongodb.com/cloud/trust
- Ubuntu security - https://ubuntu.com/security
- NPM security policies - https://www.npmjs.com/policies/security
- NodeJS security - https://nodejs.org/en/security/