Automated Privacy Compliance Auditing as a Service (DRAFT)

Benjamin J. Anderson

University of Wisconsin - Stephens Point

Stephens Point, Wisconsin

A Thesis Submitted in Fulfillment of the Requirements for the Degree Of

Master Of Science

in Data Science

December 2020

For my wife and son.

Contents

1 Executive Summary

TODO in Chapter 3.

2 General Company Description

Cereus is a software-as-a-service company offering privacy compliance auditing tools and solutions for businesses that collect customer information through their websites.

2.1 Unique Value Proposition

Consent involves more than just cookies on your website. It involves all information you share with your vendors and partners including network traffic. Cereus actively scans, monitors, and analyzes your website to identify compliance infringements in accordance with your business rules.

2.2 Mission Statement

To remove privacy barriers between companies and their partners, and enable them to communicate in an efficient, compliant way.

2.3 Vision Statement

Proactively detect all privacy compliance infringements defined by our clients before they reach their customers.

2.4 Values Statement

- Reliability
- Transparency
- Growth

2.5 Company Goals and Objectives

Our primary goal is to establish ourselves as a thriving company that leads the privacy industry by providing automated, insightful, audits to ensure that companies and their partners are sharing customer information in accordance to their business rules.

2.6 Business Philosophy

Some companies spend millions to establish a privacy program. Everyone else gets Cereus.

2.7 Industry Overview

The privacy technology industry is a rapidly growing field. In 2020, it is the fastest growing technology sector that includes the fastest growing company in the U.S (Hughes, 2020). As more consumers become impacted by massive data breaches in which sensitive, personally identifying information (PII), is exposed; consumer awareness and the call for data processing regulations are expected to be on the rise.

There are already governmental regulations impacting businesses in the U.S. The Health Insurance Portability and Accountability Act (HIPAA) highly regulates patient information and how it is stored (Centers for Disease Control and Prevention, 2018). The California Privacy Protection Act (CCPA) regulates the selling of user data collected by a business for consumers in the state of California (California Legislature, 2018). The Children's Online Privacy Protection Act (COPPA) imposes requirements on website operators on collecting information from children under the age of 13 years old (FTC, 1993). Lastly, the General Data Protection Regulation (GDPR) gives European Union citizens the right to manage their information any business has collected on them and requires explicit consent before information is collected (Council of the European Union, 2018).

With all these regulations, foreign and domestic, companies that collect information from their customers are relying on the assistance of privacy technologies to operate within the bounds of new regulations and meet consumer privacy expectations (Meehan, 2019). This has attracted massive funding, and, in July of 2019, OneTrust raised $200 million in a Series A investment, TrustArc raised $70 million Series D, Privitar raised $40 million series B, and BigID raised $30 million Series B (Wood, 2019).

In four years since its founding, as of August, 2020, OneTrust is valued at $2.7 billion (Hughes, 2020).

2.8 Market Segment Overview

Any business that operates a website and collects data and analytics on their customers is subject to the regulations in which the customer originates. This also applies to the jurisdiction in which the business operates. Cereus can provide auditing for companies with a single website, to large enterprises with hundreds. It may prove difficult for small businesses with a single website to justify the expense of privacy audits when they are not often the subject of privacy lawsuits (LaNou, 2020). Our primary focus will be medium to large organizations maintaining multiple websites.

2.9 Legal Form of Ownership

Cereus will be organized as a small business corporation (S corporation). As an S corporation, we will be allowed to collect funding and pass off any business profits and expenses to its shareholders without the additional taxes applied to C corporations.

2.10 Guiding Principles

The "Living Principles for Design" framework (Hamlett, 2020) was applied to outline how Cerues can maintain a sustainable design while achieving our objectives along the following dimensions:

2.10.1 Environment

The direct environmental impact of Cereus is expected to be minimal. We will provide software-as-a-service (SAAS) and will rely on cloud services to manage our operations. Cloud services, such as Amazon Web Services (AWS), are composed of large computer networks in which infrastructure is shared with other AWS customers (Amazon, 2020). Our physical hardware is limited to the machines required to manage services running on the cloud platforms.

With Cereus's services operating in the cloud, a central office space for employees is not required and will further reduce the company's environmental impact.

2.10.2 People

The societal impact of our company and its services are restricted to the transparency of the companies that use it. We offer detailed reports from our audits that can provide insights into how customer information is shared between a website and its partners. If companies choose to share these reports, their customers will better understand how their information is used in exchange for the services the company provides. This has the potential to improve the relationship between a company and their customers – possibly making them more apt to sharing personally identifying information.

2.10.3 Economy

Cereus's operations are expected to reduce the amount of time required to conduct a compliance audit against websites. These actions will minimize the manual auditing cost and likelihood of our customers being subject to privacy lawsuits. Our customers can then focus and dedicate more resources towards achieving their goals and growing their business. Our

overall economic impact is limited to the the actions of our customers and is expected to be minimal.

2.10.4 Culture

Cereus has the potential to influence organizations to be more transparent about the sharing of information on their customers with their partners. Traditionally, data processing and sharing are often confidential and kept internal; but with privacy becoming a concern for consumer – transparency will soon be an expectation (Meehan, 2019).

3 Products and Services

Cereus aims to become the leader in cutting-edge privacy compliance auditing tools. These tools will assist our customers to quickly and efficiently identify privacy and compliance issues on their websites. Most of our products will be offered as software-as-a-service with additional professional services also being made available.

3.1 Description of Products and Services

Our products and services emphasize the privacy auditing process outlined by the Information Systems Audit and Control Association (ISACA). After our customers integrate with Cereus, they will be able to conduct audits with the following steps (ISACA, 2014):

- Establish Context

  There is no universally agreed-upon understanding of privacy and interpretations differ from country, culture, or organization. Within our systems, you will be able to define the context in which certain information you collect can be used.

- Identify privacy risk

  Our recommendation engine will assist our customer with identifing any emerging risk areas when allowing a component of your website to run in a certain context. Some emerging risk areas include: governmental regulations, big data, mobile applications, social media, and cloud security.

- Analyze privacy risk

  Cereus will assign a risk score associated with a component of your website should it load under certain contexts. Our customers can use this risk score to evaluate or curate any control measures, such as a privacy policy, to help mitigate any associated risk.

- Evaluate privacy risk

  After evaluating the associated risk score and your company's control measures to mitigate any associated risk, Cereus can suggest the remaining, residual, risk score.

- Manage privacy risk

  Our customers will then determine the steps to take for risk management. Such measures include avoidance, transfer or reduction to an acceptable level, and cost vs. benefit of the risk treatment. Consent management providers can assist our customers with privacy risk management if needed.

- Communicate and consult

  Audits can be scheduled based on your company's requirements. The audit report can be exported and provided to stakeholders to quickly address any areas of concern.

- Monitor and review

  To evaluate the performance of your risk management solution, active monitoring and periodic reviews of your privacy management and risk mitigation implementation is required. Governmental regulations and internal processes may change, which may impact your privacy risk management practices. Our automated auditing process will assist our customers ensure that their risk management processes are based on up-to-date information.

3.1.1 Compliance Auditing

The Cereus compliance auditing system is a series of rules and processes configured by the user to ensure their websites meet the compliance standards that they have defined. It can be configured to scan websites during the businesses development cycle, automatically through the API services, or manually through the our user interface. There are five main components to the compliance auditing system: the confirmation system, crawler, setup wizard, rules engine, report, and recommendation engine.

3.1.1.1 Confirmation System

The Cereus confirmation system, Figure 3.1, prevents organizations from conducting audits on domains that they do not own or manage. This ensures that our customers cannot use the auditor to evaluate their competitors websites and privacy practices. When a new property (website) is added through the user interface, the user will be guided through a series of processes to confirm ownership of the domain before allowing an audit to be conducted. This process is dependent on the plan the customer is subscribed to. Any attempt to scan an unconfirmed domain will be rejected.

The free tier is a highly restricted plan that limits the capabilities of the auditor. This tier is intended for small business owners running a single website with limited content. For the initial confirmation of ownership of the domain, free tier users will be required to own an email address associated with the domain they are requesting to audit. In many instances, an email such as "webadmin@example.com" are dedicated to the management of the domain. Once the customer creates the property in our user interface, they can then request a validation email be sent to their inbox with a confirmation link. When confirmed, the customer will be provided an HTML metadata tag to be included on the pages they would like to be scanned.

In the event the property on the free tier expires or is sold, Cereus will no longer be able to audit the website due to the metadata tags not being present on the site.

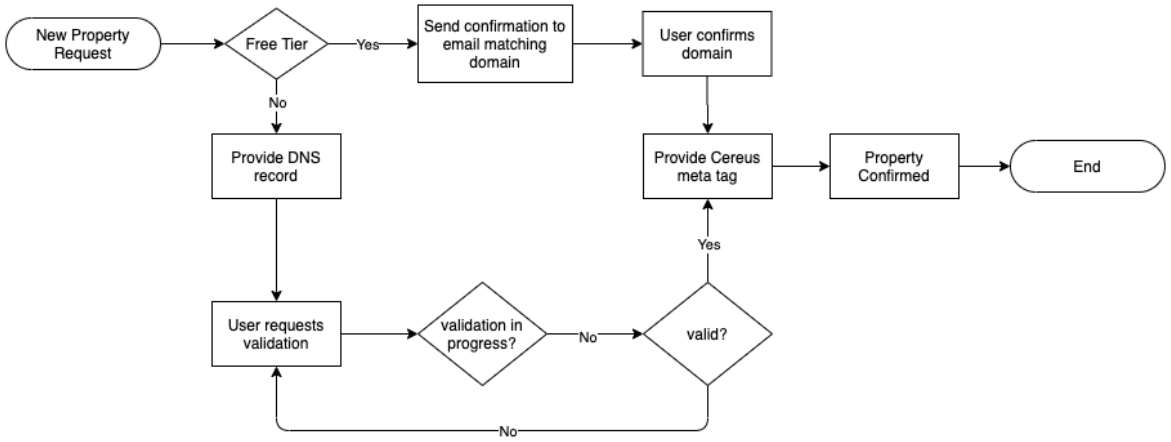

Figure 3.1: Property confirmation system - The Cereus website confirmation system to ensure the organization owns the domain prior to scanning.

All additional tiers offer unrestricted access to our services available in the tier. Our services can provide insights into the privacy operations of the company, so additional confirmation steps are required. To confirm ownership of the property, a TXT DNS record defined by our systems will be provided to the client. In the event a domain is acquired by a new party, through the sale or expiration of the domain, we will lock access to previous reports and disable auditing services when the TXT DNS record is no longer present. The initial client who set up the domain in Cereus will be notified of the change and can re-validate if the DNS record was removed by mistake.

3.1.1.2 Crawler

The Cereus crawler is a highly intelligent and configurable service that uses artificial intelligence to scan, monitor and classify information from validated websites. This includes all network traffic, cookies set on the page, call-trace information, performance data, request type, response status codes, headers, and redirect flags. This data is then processed by our rules validator and aggregated by the report generator. Our recommendation engine will then suggest what actions should or must be taken to bring the website back into compliance. Unlike our competitors, this includes the exact line and position of code where the non-compliance occurred.

Configuration

The crawler can be configured in a variety of ways to meeting an organization's auditing and compliance needs. From a set or random schedule, organizations can ensure their publicly accessible website are scanned on regular basis to ensure compliance and to verify the websites are only updated through approved channels. Once the data has been processed by the rules validator and the report has been generated, the customer will receive an email notification with the audit report. Audits may also be conducted manually by a user through the Cereus user interface or triggered through the API services.

Our customers may also specify a geographic location for which the crawler will originate from. Depending on the geographic location of the visitor, privacy regulations can differ and the company may apply a different set of business rules. Running the crawler in a targeted geographic location allows our customers to validate their website's behaviors.

Actions

In many instances, features of a website may require an authorized user or action to be taken in order for functionality to trigger. Our customers can specify actions the crawler can take in order to trigger the specified site functionality. This includes clicking the specified components, scrolling, or waiting for an event to trigger.

Data Extraction and Transformation

The network and cookie data are extracted in a semi-structured format that can be translated to a flat SQL tables for processing by the report generator. Data received by the proxy server can be consolidated with debug information sent by the browser. Figure 3.2 outlines the data collection and transformation process.

The crawler captures the requested URL for users to construct rules on the domains, locations, and query parameter. The requested URL is broken down into the requested protocol, domain, path, query, and location hash. This fragmentation significantly reduces the amount of processing required by the rules and recommendation engines. It also allows our customers to optimize their reports by establishing aggregation or exclusion rules based on part of the URL. In the instance data is sent to the server, for example, in a POST request: the data is converted to a HashMap and stored as JSON in the metadata column.

All request and response header information, depending on the browser meta event, is formatted as JSON and stored in the headers column. Signals such as the Do Not Track header signal (DNT) can be sent from the browser to indicate that the user would prefer privacy rather than personalized content (Mozilla, 2020). Site partners may respond to the

DNT header, or some other setting, that the customer can monitor with the Cereus rules engine.

3.1.1.3 Setup Wizard

Upon the initial implementation of a property within our system, Cereus will prompt our customers asking if they would like to set up some default rules prior to the initial audit. The user will identify the industry in which the property belongs to and select some common partners often affiliated with the industry. At the end of the setup process, Cereus will apply the rules to the property. This is an optional process and the generated rules can be modified or removed if needed.

3.1.1.4 Rules Engine

Our customers have the option to establish rules associated with a network request or cookie to determine whether or not they meet compliance standards. These rules can be configured to target specific geographic location to determine if an entire URL, domain, protocol, query path, or parameter, based on the specified condition, meets compliance expectations (Figure 3.3).

Multiple conditions may be applied to a single URL or cookie as an OR conditional. Rules may also be grouped to establish an AND conditional between two or more rules. Cereus's rules engine supports matches for values that: contain a value, equal a value, does not contain a value, does not equal a value, or whether or not the value matches a regular expression. Based on specified conditions, the user can specify whether or not to flag the request as compliant or not.

These rules, by default, are set at an organizational scope. All properties under the organization, when the rules engine processes a crawl, will have the same rules applied. Rules may also be overrode at a property level when exclusions are needed.

3.1.1.5 Reporting

Once processed by our proprietary Artificial Intelligence and patent-pending algorithms, the Cereus audit provides insights into the requests made on the site. This incorporates the response status code, the type, size of the information exchanged, the amount of time for the request to complete, and whether or not the request met compliance expectations (Figure 3.4). Any information that has no data or rules associated with it appears as a question mark (?) icon.

Each row can be broken down to dive into the information associated with the request. The deep dive includes the headers associated with the request, a cache of the original rules

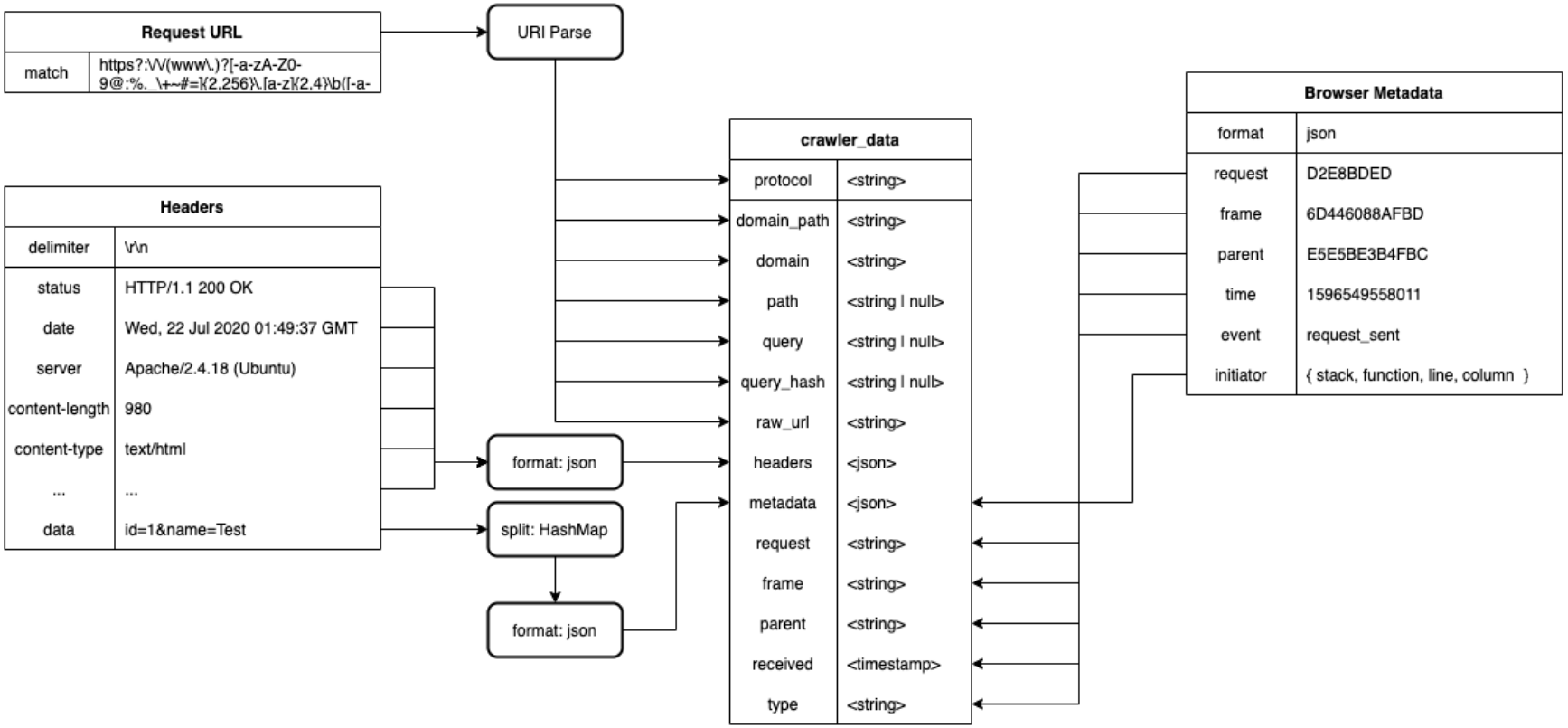

Figure 3.2: Crawler data transformation - The extraction and transformation of data received by the Cereus crawler.

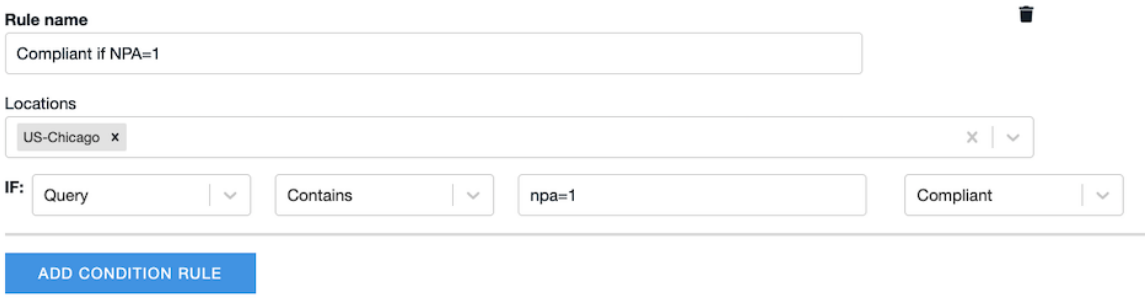

Figure 3.3: Rules definition interface - The Cereus rules engine can be configured to determine whether or not a request meets compliance standards based on the conditions specified by the user.

associated with the request (and their validation status), query parameters, data sent to the server, cookies, and the initialization chain. Web administrators and compliance managers will be able to quickly reference the audit report and identify where the compliance infringement originated.

The report filters can be used to dynamically query information from it. Users can check for specific URLs, whether or not requests were compliant, the associated category with a URL, or the page in which the information was found. Our customers will receive a notification when the audit identifies requests out of compliance and can use the filter functionality to quickly pull up the request and rules information.

3.1.1.6 Recommendation Engine

Cereus can make suggestions for requests and cookies that have yet to be classified by the organization. This classification system is powered by the categorizations of requests and cookies by other organizations. The system will also crawl the domains, paths, and cookie hosts to grab meta information to improve the accuracy of the recommendations. A risk score will also be assigned to the requests and cookies based on whether or not other organizations have flagged it as necessary to their operations.

Risk Score

When a new request or cookie is identified on a website, through an initial or later scan, an associated risk score with allowing it to load in a list of geographic locations. This score is merely a suggestion based on the operations of other organizations and no action needs to be taken.

$$s_{risk} = \frac{r_n}{O_r}$$

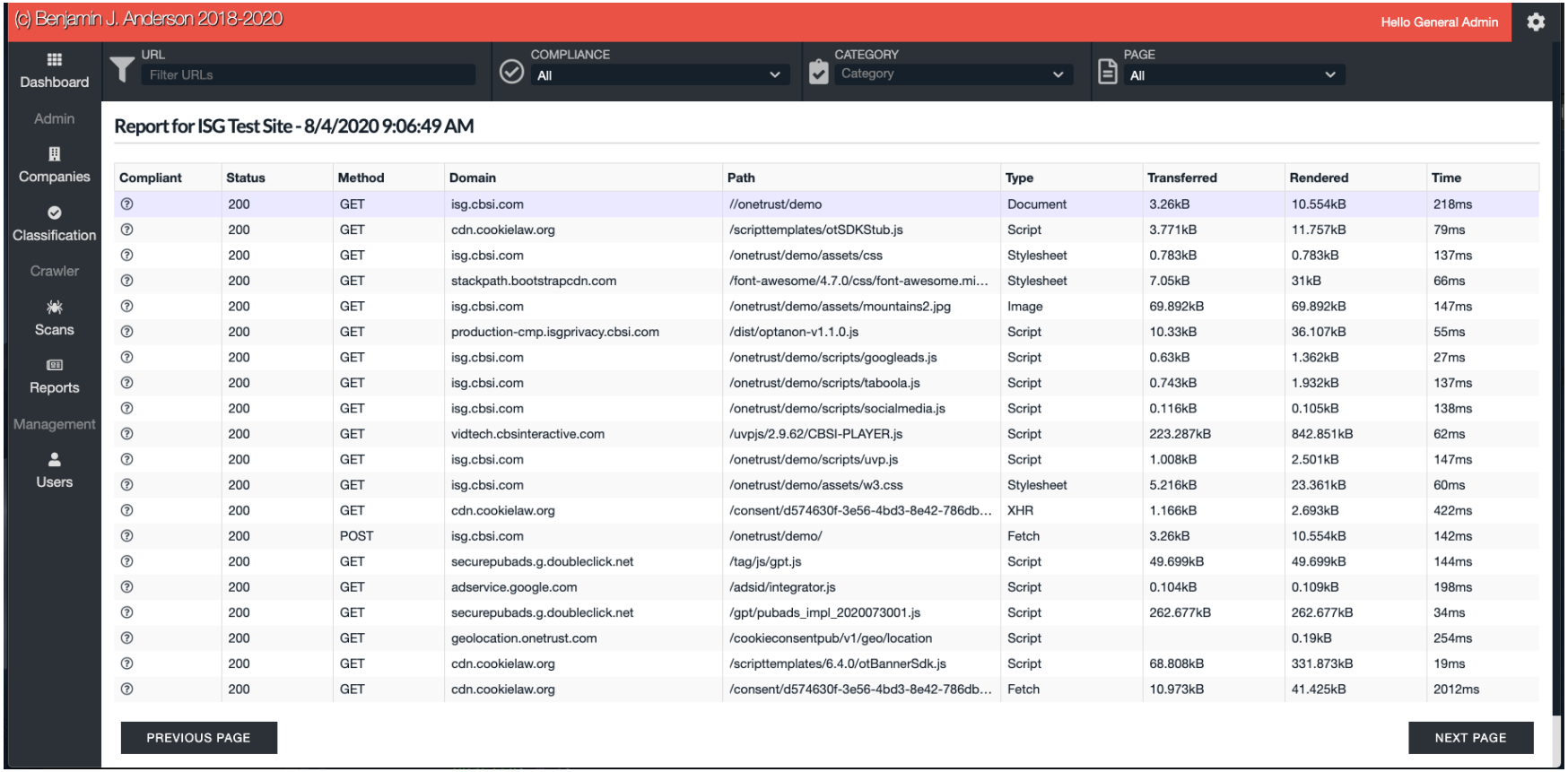

Figure 3.4: Cereus audit report - The Cereus audit report provides clean, familiar, interface for privacy compliance teams.

The risk score is computed as the number of records classified as necessary r_n divided by the number of organizations in which r_n is found (O_r). This will always result in a ratio between 0 and 1 in which intervals of $\frac{1}{3}$ will determine if the request will be rated as: low, intermediate, or high risk.

To reduce the possibility of organizations incorrectly flagging a request due to the scoring system, the risk score will only be provided when a sample size of at least 20 organizations have classified the request or cookie.

Categorization Recommendation

Cereus will provide categorization recommendations for requests and cookies based on meta information extracted from the request or cookie's origin. The recommendation engine will also incorporate organization classification information pertaining to the request or cookie. Classification within Cereus's internal systems are expected to be single words or small phrases, much like meta tag information present on a web page. This information can best be represented as a bag of words. There's no context to meta tag data or the collection of classifications entered by users, so the representation of language or order has no meaning (Manning, 2008).

The recommendation engine is powered by a Bernoulli document model. The model takes a document and partitions it into a feature vector of binary elements. If a word is found in the document, it will receive a value of 1, otherwise 0. It does not take into account the frequency of a word, but whether or not the word is present. The model can then calculate the probability of a word occurring in a document with a specific classification, as well as taking into account the probability of it not occurring.

To save reduce the computational power required to provide recommendations, the model calculates estimated probabilities for a request or cookie belonging to a category.

$\hat{P}(w_t|C_k)$ defines the estimated probability that the word w_i occurs in a document, D, with the classification (C) k. The estimated probability that the word w_i not occurring is $1-\hat{P}(w_i|C_k)$. V represents the model's vocabulary and v consists of the feature vector of the document. The product of the probability of each item (i) in the feature vector occurring or not occurring determines the overall estimated probability of the document being classified as class k ($\hat{P}(v|C_k)$).

$$\hat{P}(D|C_k) = \prod_{i=1}^{V}[v_i\hat{P}(w_i|C_k) + (1-v_i)(1-\hat{P}(w_i|C_k))]$$

There are two parameters for this model: the probabilities of each word in the document class ($\hat{P}(w_i|C_k)$) and its prior probabilities $P(C_k)$. The estimated probability that a word

w_i occurs in a document is the number of documents n classified as k divided by the total number of documents N classified as k.

$$\hat{P}(w_i|C_k) = \frac{n_k(w_i)}{N_k}$$

Where the prior probability of class k can be estimated as the relative frequency of documents containing class k.

$$\hat{P}(C_k) = \frac{N_k}{N}$$

The output of the model will be an array of estimated probabilities. Each probability in the array is associated with a category defined within Cereus's system. The category with the highest probability is recommended to the user.

3.1.2 Notifications and Alarms

Users can configure Cereus to notify the specified stakeholders in the event that an audit discovers new requests or cookies, compliance checks fail, or when a new scan is scheduled. To reduce the likelihood of users ignoring these notifications, they can be configured on an organizational level or per property. These settings ensure that notifications are sent only to the significant parties and reduce email clutter.

3.1.3 API Services

Professional and Enterprise customers may use Cereus's application programming interface (API) services to automate their privacy operations. Our customers who actively curate new content, use a continuous integration (CI) system to control website deployments, or those who may not follow a strict deployment schedule can use the API to integrate Cereus into their daily business processes.

Crawl Requests

The crawl request endpoint allows customers to trigger a crawl against a property. By default, this will scan all paths defined on the property. Our customers also have the option to specify the path(s) they would like to crawl in the event they want to target only new or updated pages.

Web Hooks

When an audit has completed, it will notify stakeholders stating whether or not compliance checks were successful. Cereus can also notify other computer systems. Through web hooks, our customers are able to configure Cereus to send audit status messages to their computer systems. The customer can then process the message and take action against the messages.

3.1.4 Professional Services

Our professional services will serve a clientele requiring guidance on privacy regulations applicable to their operations. We will consult the client to identify the best plan for the client and assist with the configuration of our tools to reflect their needs. Training services for Cereus's products will also be offered.

3.2 Competitive Advantages and Disadvantages

The privacy industry, though relatively new, already consists of some large organizations servicing a significant amount of the market share. In April of 2020, OneTrust had a 35.5% stake in the data privacy management software market (OneTrust, 2020). We intend to enter the privacy software market by automating an otherwise manual process: compliance auditing.

3.2.1 Competitive Advantages

1. **Founders are experienced software engineers**

   Cereus's founders are experienced software engineers. Providing software-as-a-service requires the software to be developed, maintained, and readily available. Cereus's founders will be able to construct its systems to be scalable and maintainable to accompany organizations of any size.

2. **Founders are experienced product managers**

   Cereus's founders are experienced product managers. The founders are able to clearly define Cereus's product vision and prioritize feature requests to meet customer requirements.

3. **Founders have a background in data science**

   Cereus's founders have a background in data science. Cereus will collect, process, and store large amounts of information for their customers. When large datasets come into

play, storage costs, data integrity and performance issues are common. With a formal background in data science, Cereus's founders are able to mitigate all of these problems.

4. **Cereus provides actor stack traces**

   In the event a request or cookie fails to meet company compliance standards, Cereus provides stack trace information that allows the user to locate the file and exact line of code that triggered the validation failure. This saves the company time and money from having to manually trace the source.

5. **Cereus extends cookie compliance**

   In addition the industry standard of cookie compliance (user consent is required only for cookie tracking), Cereus also provides insights into network requests. This provides a full-scope audit of the website and how information is shared between it and its partners.

3.2.2 Competitive Disadvantages

1. **Cereus is not a consent management platform**

   Cereus does not provide consent management services for its customers; it augments existing consent management platforms and solutions. Cereus's customers will need to purchase a consent management solution or implement their own.

2. **Adoption requires legal support**

   Tools used to automate legal compliance often require privacy attorneys to validate whether or not the tools are effective (Merken, 2019). Cereus will require legal support in order for companies to adopt its services.

3.3 Pricing Structure

We aim to serve medium to large-sized organizations managing multiple websites. We do, however, offer a flexible pricing structure, including a free tier, to accommodate companies of any size. Table 3.1 provides an overview of each tier and services provided.

3.3.1 Free

We offer our auditing services, with limitations, for free. Free-tier users are authorized to register one website but will not be able to benefit from the company's API, web hook, recommendation, notification, or professional services. They are limited to running 10 audits per year against a running total of 12 pages.

3.3.2 Personal

The Personal tier enhances the Free tier by incorporating our notification system into the audit process. It expands auditing to five pages of the customer's website once per month. The recommendation engine will further assist our customers by offering insights into the categorization of requests and cookies found in the audit. One user may be set up to access the organization's information within Cereus.

This is a monthly subscription that can be canceled at any time.

3.3.3 Professional

The Professional tier provisions up to five users to be set up to access the organization's information within our systems. It expands the Personal tier to allow the our customers to register up to five of their websites. Each website can have up to ten pages audited twice a month with the option to integrate web hooks into their auditing processes. Our professional services are also offered for an additional fee.

This is a monthly subscription that can be canceled at any time. A yearly subscription is also available at a 5% discount.

3.3.4 Enterprise

Enterprise customers have access to all of our services in addition to 40 hours of professional services per year. The option to have the auditor crawl a sitemap as opposed to manually defining pages, with a limit of 100 pages, is available. 25 websites can be registered and audited up to 4 times a month. 25 users may be set up to access the organization's information.

This is a yearly subscription that cannot be canceled.

3.3.5 Select

There is no one-size fits all pricing model that fits every customer's needs. Our Select tier offers customers the ability to tailor a customized experience at a fixed, yearly, rate.

3.4 Industry Background

With new governmental regulations, data breaches, fines, and growing awareness of how personal data is collected and processed, the privacy technology industry is growing rapidly – with one company being valued at over $2.7 billion after being founded only 4 years since its valuation (Hughes, 2020). With rising successes of privacy software vendors, we also see large data collection and processing companies being hit with record-breaking fines. In 2019,

Tier	Subscription	Price	Users	Websites	Pages	Scans/Mo	Prof.Services	API	Web hooks	Notifi.	Rec.	
Free	Monthly	$0.00	1	1	1	1	✗	✗	✗	✗	✗	
Personal	Monthly	TODO	1	1	5	1	✗	✗	✗	✓	✓	
Professional	Monthly	TODO	5	5	25	2	Additional	✗	✓	✓	✓	
Enterprise	Yearly	TODO	25	25	100	4	✓	✓	✓	✓	✓	
Select	Yearly	Custom	N	N	N	N	Opt.	Opt.	Opt.	Opt.	Opt.	

Table 3.1: Cereus pricing tier and services provided.

the FTC fined Facebook $5 billion and imposed strict privacy restrictions (FTC, 2019). This set a new precedent that all companies are accountable for the decisions they make about their users' privacy.

As customers are becoming more aware and passionate about what companies are doing with their data, the need for privacy technology is expected to continue to grow (Meehan, 2019). Governmental regulations will continue to focus on making companies accountable for the data they collect and require them to follow various guidelines. In instances where the regulations are too complicated for companies to implement or manage on their own, privacy vendors can step in and assist.

3.5 Target Market Segment

At Cereus, we offer flexible subscriptions for any sized organization that's looking for privacy compliance auditing. It may, however, be difficult to market smaller companies that our services are worth mitigating the risk of being fined. We will focus our marketing efforts towards medium-to-large organizations, with multiple websites, who are currently, or will be, working with a consent management system.

Medium-to-large organizations will often attract a larger, more diverse, customer base. Their customers determine the privacy regulations the company has to adhere to. As a company grows, collects, and process more information, it may attract the attention of customers or governmental agencies – making the company more subject to fines or legal action should they be found out of compliance. With more regulations, our customers would have to dedicate resources towards manually auditing their websites for each regulation to mitigate the risk of lawsuits or fines. At Cereus, we automate this process and follow the rules defined by our customers without any margin of error, making our costs more justifiable in contrast to smaller companies.

4 Marketing Plan

4.1 Industry Background

4.1.1 Market Size

According to QY Research, North America and Europe are currently holding the largest share for the privacy management software market (QY Research, 2020). In 2019, this market size was estimated at a US $808.8 million and estimated to reach a US $6.1 billion by the end of 2026 with a compound annual growth rate of 33.1% from 2021 through 2026. These estimates were based on privacy management software trends in North America, Europe, China, Southeast Asia, India, and Central and South America.

4.1.2 Market Trends

With the passing of the General Data Protection Regulation (GDPR) for European Union citizens and large data breaches exposing the personal information of hundreds of millions of consumers, privacy awareness has led to the passing of privacy regulations across an increasing amount of countries (PrivacyPolicies, 2019). Other countries, including Brazil, Canada, Australia, and various U.S. states have enacted, or began enforcing, their privacy laws with heftier fines or sanctions. The primary difference between regulations being passed now, in contrast to the past, are the heavy fines and sanctions regulators can impose on companies who are non-compliant (Matteson, 2020). These sanctions and fines are designed to force businesses to comply with regulations and how they process their customers information.

In a survey conducted by Auxier, Rainie, Anderson, Perrin, Kumar, and Tuner, 62% of Americans believe it is not possible to go through daily life without having their data collected, 81% believe that they have very little to no control over their information companies collect about them, and 72% feel that all, or almost all of what they do online is being tracked (Auxier, 2019). The survey also revealed that most Americans are not confident that companies would publicly admit to misusing their customers data, but still say data collection and usage is acceptable for processing in some ways.

Citizens of countries belonging to the European Union, 1 year since the passing of GDPR, are rapidly becoming aware of their privacy rights (European Commission, 2019). A survey conducted by the European Union Commission against 27,000 Europeans revealed that 73% of citizens have heard of at least one of their privacy rights with the highest awareness being the right to access their own data (65%), the right to correct their information a company has on them (61%), and the right to opt-out of direct marketing (59%).

Other trends involve the usage of smart mobile devices. According to the FTC, information from smart phones, such as location information, is considered sensitive (Federal Trade Commission, 2012). The location information can be used to provide, unwanted, targeted marketing to consumers based on their movements. Mobile applications can use the hardware identifiers associated with devices to uniquely identify consumers or potentially access utilities no required for the intended purpose of the application (Tama, 2012).

4.1.3 Growth Potential and Opportunity

In section 4.1.1, QY Research identified a compound annual growth rate of 33.1% from 2021 through 2026. Section 4.1.2 identified emerging trends such as: privacy awareness, information control, and concerns with smart mobile devices. With a rapidly growing market and rising awareness and concerns with privacy, we can anticipate that governments will further regulate how companies can process consumer information. With further regulations, the demand for consent management vendors will grow (as predicted by QY Research) and automated privacy auditing services to ensure privacy risk is mitigated by the vendors will be in similar demand.

4.1.4 Market Barriers

Even with governmental regulations, hefty fines, and sanctions; adoption of consent management platforms (CMP) isn't widespread and integrations often don't meet compliance regulations (Nouwens, Liccardi, Veale, Karger, & Kagal, 2020). In an audit conduced by Aarhus, Cambridge, and UCL Universities, of the top 10,000 websites in the UK, only 11.8% met the minimal requirements set by their audit based on European law. Of this sample, only 20.35% of websites report to use a CMP. In addition to the top 10,000 websites, in a sample of 1,000 consent management platform vendors, 95.8% provided either no consent choice or confirmation only.

A long-term study of the impact of GDPR on cookie placement also revealed that (Trevisan, Traverso, Bassi, & Mellia, 2019):

1. 49% of websites placed cookies before receiving consent.
2. 28% of websites didn't provide any consent mechanism.
3. The percentage of websites violating GDPR stayed constant over the course of 4 years, indicating any consent mechanisms implemented were ineffective.

The primary barrier for Cereus to successfully provide privacy auditing services will be industry acceptance. With many vendors offering consent management solutions, a majority

of integrations failed to meet compliance standards (Nouwens et al., 2020). We will have to differentiate ourselves from consent management providers looking for a piece of a booming market and emphasize that we are not a consent management solution. Cereus is an auditing and risk mitigation solution that evaluates integrations with CMPs. Should a CMP integration fail to adhere to the defined rules within our auditing system, those failures will be reflected in the audit report.

4.1.5 Market Changes

As an automated privacy auditing company, Cereus is moderately affected by fluctuations in the privacy software market due to changes in privacy legislation and regulation. Response to changes will, generally, be the responsibility of our customers – they will be able to alter their business rules within our systems to ensure that any new legislation is accounted for. To support our customers, our professional services staff will receive training on any new or changes to legislation.

In the event legislation expands or is implemented in geographic locations that we do not yet support, our Engineering staff will make the appropriate changes prior to the enforcement of the legislation.

4.2 Products and Services

Customers with a privacy program may see auditing as the next logical step towards ensuring compliance and privacy risk mitigation measures are being followed across all of their properties. Depending on size of their organization, manual auditing may be a possibility. Customers will also have to consider the costs associated with manual auditing, the possibility for human error, and the frequency in which the audits must be conducted.

Other customers looking for privacy management solutions may not consider auditing and risk management as the first priority. They understand that their websites need to comply with relevant laws and a consent management platform can assist them with compliance. These customers will look through our products and services and will quickly identify that we do not offer a consent management solution, so why would they choose our services?

We, Cereus, agree that a consent management platform seems like the first logical step when viewing privacy legislation trends and actions your competitors may be taking. We also believe that a consent management platform may not be required given your audience, business practices, and tooling. After conducting an audit and identifying any privacy risks your businesses are susceptible to, you will be able to determine an appropriate course of

action. That course may include implementing a consent management platform or possibly using open-source solutions that will mitigate risk.

With the highly customizable products and flexible pricing (even free) we offer, customers will see the potential to: identify whether or not they need a consent management platform, the associated risk of not implementing one, and the effectiveness of their integration should they implement a consent management platform.

4.2.1 Auditor and Reporting

Any customer, whether they have implemented a consent management platform or not, will first look towards how auditing and reporting will be provided. More importantly, what information the audit contains and how it can be distributed.

Unique Selling Proposition

Define your compliance rules and actively monitor all your properties to ensure that they meet your compliance standards.

Features and Benefits

- Customizable

  Define your compliance rules within our system for our proprietary crawler to validate against your websites.

- Interactive

  Verify the results of your audit in our interactive report. Identify which of your properties failed to meet your compliance rules and where the infringement occurred on the website – down to the line of code causing the infringement.

- Exportable

  Export your audit into a distributable format to share with your stakeholders.

- Automated

  Schedule audits to reflect your development cycle and quickly identify compliance infringements before they are distributed to your customers.

4.2.2 Notifications and Alarms

Customers may wonder how long an audit will take to run against their websites and, unfortunately, there is no definitive answer. Audits depend on how many pages the crawler must scan and the execution speed of your website. All of our customers, except those using the free tier, may specify who, if anyone, will receive notifications of when audits complete for each property, as well as an alarm should compliance checks fail.

Unique Selling Proposition

Allow Cereus to grant you peace of mind by notifying you should one of your websites fall out of compliance.

Features and Benefits

- Customizable

  Specify which stakeholders, at an organization or property scope, to notify when audits complete, and compliance checks fail for any of your websites.

- Prompt

  No need to step away from your task to check on the status of your audits, allow Cereus to notify you as soon as your audit is complete.

4.2.3 Classification and Recommendation Engine

All our customers, excluding those on the free tier, will see our audit system provide a risk score, residual risk score, and suggest a category for partners running on your website that you have yet to review. Many will ask why such functionality is provided as their legal department will determine the rules associated with a partner. To assist with the evaluation of partners, we've aggregated resources so you don't have to.

Unique Selling Proposition

Allow us to assist you with establishing your business rules by providing insights into the operations of your partners and suggest a course of action for your websites.

Features and Benefits

- Automated

Cereus will automatically calculate the risk score and residual risk score should a partner load on your site in specific geographic locations. This score is based on the actions taken by our other customers, vendor information, and your privacy policy.

- Productivity

  Make informed decisions about mitigating any risk associated with your partners through the resources we've collected pertaining to their operations.

4.2.4 API

Organizations actively developing their websites will look for functionality that enables them to integrate privacy compliance and risk management into their daily operations. Our application programming interface (API) provides a variety of features to support a continuous integration system. This is available for our professional, enterprise, and select customers only.

Unique Selling Proposition

Automate privacy compliance and risk management by integrating our API into your business operations.

Features and Benefits

- Efficiency

  Include privacy compliance and risk management in your continuous integration system. Run audits on-demand and let us send the results to your systems.

4.2.5 Professional Services

Not all organizations have the time or resources to dedicate towards running a privacy audit and risk management division. Professional and enterprise customers have the option to allow our professional services staff configure our systems to meet your privacy compliance and risk management requirements.

Unique Selling Proposition

Allow us to assist your organization by configuring our systems to meet your privacy compliance and risk management requirements.

Features and Benefits

- Support

  Staff readily available to assist with your integrations, setup, and general questions.

- Active Management

  Our professional services staff can configure and manage our systems to meet your privacy compliance and risk management requirements.

4.3 Customers

Though we offer flexible subscriptions for any sized organization that's looking for privacy compliance auditing. It may, however, be difficult to market smaller companies that our services are worth mitigating the risk of being fined. Our customers will primarily be medium-to-large organizations, with multiple websites, who are currently, or will be, working with a consent management system.

4.3.1 Demographic Considerations

The geographic location of our customers and their client-base are some factors to consider when defining the demographics of our customers. The location of our customers and their client-base define the regulations they have to adhere to. Should an organization chooses to restrict their client-base due to data and privacy regulations, our services may not be required.

4.4 Proposed Location

Cereus will be incorporated in the United States. There will be no leased or physically owned real estate. Correspondence pertaining to Cereus will operate through a PO Box.

4.5 Pricing and Positioning Strategy

TODO Chapter 2

4.5.1 Operational Costs

Operational costs are broken into two categories: infrastructure and administration. Infrastructure includes all technology requirements to support operations. Administration includes expenses pertaining to wages, marketing, accounting, and professional services.

4.5.1.1 Infrastructure Costs

Cereus will manage its infrastructure through Amazon Web Services (AWS). AWS offers an alternative solution to physical infrastructure management at a competitive rate. For minimum operations, Cereus will require the following services from Amazon:

- **1 A1.Medium Spot Instance**

  The scan servers to collect information from our customer's websites can be reserved spot instances that operate only on demand. These servers come at a discounted rate, but introduce some latency between a customer's audit request and its completion.

  Pricing can be calculated with the following formula where t is time in hours and n_{a1s} is the number of servers:

$$Scan_{cost} = 0.0049tn_{a1s}$$

- **2 A1.Medium Reserved Instances**

  Cereus operates on a client-server model, in which the client will operate as "serverless" through S3 and a CDN. The API services for server portion must be available at all times with limited latency. This comes at an additional cost in contrast to spot instances, but will ensure our customers can request information on-demand.

  Pricing can be calculated with the following formula where t is time in hours and n_{a1r} is the number of servers:

$$Api_{cost} = 0.0255tn_{a1r}$$

- **2 S3 buckets**

  All audit information will be stored in a private Amazon S3 bucket. Cereus's landing page and user interface will be hosted through S3 as well.

  Pricing can be calculated with the following formula where g_{s3} is the size of all S3 buckets in gigabytes (for the first 50 terabytes):

$$S3_{cost} = 0.023g_{s3}$$

- **1 Cloudfront Instance Across All Edge Locations**

  Cloudfront will serve as our CDN layer for our landing page and user interface. This will significantly reduce the latency of serving content to our customers.

Pricing can be calculated with the following formula where g_{cf} is the size of all traffic in gigabytes (for the first 10 terabytes):

$$CF_{cost} = 0.085 g_{cf}$$

- **1 Network Address Translation Gateway (NAT)**

  To protect our proprietary technologies and customer's data, Cereus's services will primarily reside in a private network within AWS. The NAT will allow our internal network to communicate with the internet.

  Pricing can be calculated with the following formula where g_{nat} is the size of all traffic in gigabytes and t is the total time in hours:

$$NAT_{cost} = 0.045t + 0.045 g_{nat}$$

- **1 Replicated Postgres Instance**

  Cereus will store its customer information and configurations in a postgres database. Data will be replicated across two instances for high availability.

  Pricing can be calculated with the following formula where t is the total time in hours:

$$RDS_{cost} = 0.072t$$

- **1 T2.Small Elasticache Instance**

  Scheduled scans and notifications will operate through a queue-consumer implementation. Amazon's Elasticache service will serve as the queue.

  Pricing can be calculated with the following formula where t is the total time in hours:

$$ELC_{cost} = 0.034t$$

- **API Loadbalancer**

  Though only 1 API instance is needed for minimum operations, should that 1 API server fail, Cereus's customers will not be able to access their configurations or audits. Running 2 API instances with a Loadbalancer will prevent all operations from halting due to one server going offline.

  Pricing can be calculated with the following formula where g_{elb} is the size of all traffic in gigabytes and t is the total time in hours:

$$ELB_{cost} = 0.025t + 0.008g_{elb}$$

- **1 Route53 Hosted Zone**

  Route53 is Amazon's domain-name server resolver. Cereus's domain service provider can route requests from our domain to our servers, through Route53, hosted on Amazon.

  Pricing can be calculated with the following formula where r_m is the number of requests (per million, up to 1 billion):

$$R53_{cost} = 0.50 + 0.40r_m$$

Using the cost formulas defined for each Amazon service, we can calculate our operations costs with the following formula (simplified):

$$I_{cost} = 0.176t + 0.0049tn_{a1s} + 0.0255tn_{a1r} + 0.023g_{s3} + 0.085g_{cf} + 0.045g_{nat} + 0.008g_{elb} + 0.40r_m + 0.50$$

4.5.2 Competitive Analysis

TODO Chapter 2

4.5.3 Return on Investment

TODO Chapter 2

4.6 Sales and Distribution

TODO Chapter 2

4.6.1 Sales Strategy

TODO Chapter 2

4.6.2 Distribution Methods

TODO Chapter 2

4.6.3 Transaction Process

TODO Chapter 2

4.7 Advertising and Promotion

TODO Chapter 2

4.7.1 Advertising

TODO Chapter 2

4.7.2 Sales Promotion

TODO Chapter 2

4.7.3 Publicity

TODO Chapter 2

4.7.4 Web

TODO Chapter 2

List of Figures

List of Tables

References

Amazon. (2020). *What is aws.* Retrieved 2020-09-23, from https://aws.amazon.com/what-is-aws/

Auxier, B. (2019, 11 15). *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information.* Retrieved from https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

California Legislature. (2018, 09 24). Retrieved 2020-09-30, from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1121

Centers for Disease Control and Prevention. (2018, 09 14). Retrieved 2020-09-30, from https://www.cdc.gov/phlp/publications/topic/hipaa.html

Council of the European Union. (2018, 05 23). Retrieved 2020-09-30, from https://gdpr-info.eu/

European Commission. (2019, 06 13). *Data Protection Regulation one year on: 73% of Europeans have heard of at least one of their rights.* Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/IP_19_2956

Federal Trade Commission. (2012, 3). Protecting Consumer Privacy in an Era of Rapid Change. Retrieved from www.ftc.gov/os/2012/03/120326privacyreport.pdf

FTC. (1993, 04 27). Retrieved 2020-09-30, from https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

FTC. (2019, 07 24). *FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook.* Retrieved from https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions

Hamlett, P. (2020). *Your roadmap for sustainable design.* Retrieved 2020-09-23, from https://www.aiga.org/roadmap/

Hughes, J. T. (2020, 08 12). *Reflecting on the growth of the privacy industry.* Retrieved 2020-09-23, from https://iapp.org/news/a/reflecting-on-the-growth-of-the-privacy-industry/

ISACA. (2014, 1). Privacy Audit—Methodology and Related Considerations. *ISACA Journal*.

LaNou, C. (2020, 09 22). Personal interview.

Manning, C. D. (2008). *Introduction to information retrieval.* Cambridge University Press.

Matteson, S. (2020, 01 29). *Data privacy: Top trends to watch in 2020.* Retrieved from https://www.techrepublic.com/article/data-privacy-top-trends-to-watch-in-2020/

Meehan, M. (2019, 11 26). *Data privacy will be the most important issue in the next decade.* Retrieved 2020-09-23, from https://www.forbes.com/sites/marymeehan/2019/11/26/data-privacy-will-be-the-most-important-issue-in-the-next-decade/

Merken, S. (2019, 08 12). *Companies turning to tech vendors for privacy compliance tools.* Retrieved from https://news.bloomberglaw.com/privacy-and-data-security/companies-turning-to-tech-vendors-for-privacy-compliance-tools

Mozilla. (2020, 05 21). Retrieved 2020-09-27, from https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/DNT

Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. Retrieved from https://arxiv.org/pdf/2001.02479.pdf

OneTrust. (2020, 05 27). *Idc releases first worldwide data privacy management software market shares report.* Retrieved from https://www.onetrust.com/idc-releases-first-worldwide-data-privacy-management-software-market-shares-report/

PrivacyPolicies. (2019, 09 04). *What's Data Privacy Law In Your Country?* Retrieved from https://www.privacypolicies.com/blog/privacy-law-by-country

QY Research. (2020, 01 14). *Global Privacy Management Software Market Size, Status and Forecast 2020-2026.*

Tama, J. K. (2012, 9). Mobile Data Privacy: Snapshot of an Evolving Landscape. *Journal of Internet Law*, *16*(5).

Trevisan, M., Traverso, S., Bassi, E., & Mellia, M. (2019, 4). 4 Years of EU Cookie Law: Results and Lessons Learned. *Proceedings on Privacy Enhancing Technologies 2019*, 126–145.

Wood, N. (2019, 05 11). *New privacy tech industry attracts massive funding.* Retrieved 2020-09-23, from https://fpf.org/2019/07/11/new-privacy-tech-industry-attracts-massive-funding/