

Kunal Mehta

Abstract

This document gives a view of major data sources, and how the upcoming announcement by Google is going to impact marketing, and analytics.

Cookieless Marketing

Data Sources and the Implications

Contents

[How analytics tools create first party cookies then? 2](#_Toc88454715)

[What’s the issue? 2](#_Toc88454716)

[Anything more that we need to know? 2](#_Toc88454717)

[How does it impact the users? 3](#_Toc88454718)

[What are first party data sources? 3](#_Toc88454719)

[What is second party data? 3](#_Toc88454720)

[What is third party data? 4](#_Toc88454721)

[What are Cname cookies? 4](#_Toc88454722)

[Seems like everything is working fine, or is it? 4](#_Toc88454723)

[How does it identify Cname implementation? 5](#_Toc88454724)

[What about Facebook? 5](#_Toc88454725)

[What are data regulations? 5](#_Toc88454726)

[How would these regulations impact our data? 6](#_Toc88454727)

[What can be our solutions in this current scenario? 6](#_Toc88454728)

[How server is the issue for us? 7](#_Toc88454729)

[What’s next for us? 7](#_Toc88454730)

[Google Privacy Sandbox 8](#_Toc88454731)

[Federated user categorization & data analysis 8](#_Toc88454732)

[Understanding user interests from across site behavior 8](#_Toc88454733)

[How can we use these templates to build our analytics? 8](#_Toc88454734)

[Identity resolution in case of cookie blockers 9](#_Toc88454735)

[Current & Future Analytics arch 10](#_Toc88454736)

[Some other identity based tailored solutions 11](#_Toc88454737)

[Unified ID 2.0 11](#_Toc88454738)

[SPARROW (Secure Private Advertising Remotely Run On Webserver) 11](#_Toc88454739)

[Fatbrick ID 11](#_Toc88454740)

[What is the solution? 11](#_Toc88454741)

[Frequently Asked Questions 12](#_Toc88454742)

**Marketing Analytics in the Cookieless World**

**Cookies**: These are nothing but text files, which contain some information, for a specific period of time. They are always browser based, stored in local system. A cookie typically has following parameters:

* Information to be stored
* Expiration time (in date format)
* Domain on which cookie is created

First party cookie: These are supposed to be created by the website that you have visited Third party cookie: These are generated by websites that are different than the website that you have visited.

# How analytics tools create first party cookies then?

To identify if the cookie is first party or third party, browser tries to see the domain for which cookie is set up. If the cookie is set up on the same domain where you are visiting, these are categorized as first party, whereas if domain is different, then it will be a third party cookie. Analytics tools use this feature to set up first party cookies.

For example, in Google Analytics \_ga named cookie is created to store the Client ID, a unique user identifier allocated by Google Analytics itself.

# What’s the issue?

On Jan 14, 2020, Google has announced that it will phase out support for third party cookies in Chrome in next 2 years. Trials for the same will be starting in end of 2020.

Chrome is infact the browser which is extremely late on this decision, they are still giving 2 years, whereas Safari and Mozilla Firefox have been blocking third party cookies for a while now. Considering Chrome is typically used by 60-70% of users, therefore we can expect impact to be maximum now, but the issue has been there for some time now.

In June 2021, Google released a statement that this blockage will now come in effect in 2023. The exact date is still not provided, but it is believed that it might be somewhere in Q2-Q3 of 2023.

# Anything more that we need to know?

One of the biggest issue that we need to prepare for is ITP 2.3, that is Intelligent Tracking Prevention version 2.3, that will be launched by Apple for Safari browser soon, though there is no confirmation on the dates.

To understand what is there in ITP which is of biggest concern to us, we need to know that this is created specifically to stop organizations, AdTech/MarTech etc. to stop tracking Apple customers. ITP earlier created a rule that first party cookies will be deleted after 7 days of creation, so that MarTech tools can’t track people consistently. In ITP 2.2 they announced that this lifespan will be curtailed to just 1 day.

So, at this point of time question can be that if ITP 2.1 already addressed this issue, why Apple needed ITP 2.3? When Apple put restrictions on cookies, all these MarTech companies started using browser’s local storage to store identifiers in place of cookies. So, now with ITP 2.3, Apple will delete all non cookies based information from the browser in 7 day rolling period from the time it was created.

# How does it impact the users?

Local storage is not just used to store information about the user and preferences of the user, but also to store shopping cart, so if a user is not visiting the website in 7 days, the whole shopping cart information can be deleted, if the user is not logged in. A logged in user’s shopping cart can be loaded from the server though.

For the purpose of this document we will only limit our discussion to cookies, both first party and third party cookies and their implications for us.

# What are first party data sources?

First party data is wat companies collect from their own sources about users, their preferences, their reviews, competition, product surveys etc. They are both offline and online data sources, typically websites, mobile apps, stores, CRM, social media, surveys etc.

Since this is the data collected from their own sources, hence this is most reliable, companies are in total control of this information, can be collected for lot cheaper cost, but after new data protection regulations across the globe, we need to ensure that we do have consent in place before collecting any such information.

# What is second party data?

This is data that is typically some other organization’s first party data, which is shared with your organization, for a cost or some other contract, for you to extract insights and create your own marketing strategies. For example, you can get data from Amazon about what people are searching for, what are the products that people are buying, user demography etc, and utilize the same to refine your marketing strategies.

Reliability of this data can’t be ascertained, but it is assumed that this data can be pretty reliable. Typically it is used to understand customer segments, demography, purchase behavior etc.

# What is third party data?

This data typically comes from aggregators, who buy huge quantities of data form publishers, generally process it as well in their own ways, and then provide the same to you for a cost. Organizations utilize this data for the volume that they can get, though reliability is always under a question mark.

Similarweb is a good example of such third-party data source, which is collecting it’s data from variety of sources, from plugins, to publishers etc. and then providing this data to you at a nominal cost. This data is typically used for market research, competitor analysis, understanding business landscape in broader sense.

# What are Cname cookies?

Its short for Canonical Name, and the idea is that while setting up the cookies, there will be an alias domain, and then there will be a canonical domain. Let me give you an example here. You go to abc.com website where Adobe Analytics is implemented. In the traditional scenario, cookie used to be set by Adobe Analytics, on the domain on adobeanalytics.com (this is just an example, it used to be set on omtrdc.net, for Omniture Regional Data Center). Now this will be classified as a third party cookie, because on abc.com adobeanalytics.com cookie is being set.

Since browsers started coming up heavily on third party cookies, Adobe came up with Cname implementation. Whenever a cookie is set, it’s set up on the first party domain name, by allocating a subdomain to the analytics vendor, in our example, to adobe analytics. Cookie will now be set up on adobeanalytics.abc.com. This domain name will be reserved in a way that when someone tries to go to this domain, the person will be redirected to abc.adobeanalytics.com, that is on the domain of adobe analytics. Thereby

To implement cname, Adobe requires clients to fill up a form, and these details, along with a fee are now then used to buy the domain on the DNS, to map alias to the canonical domain. Though Adobe still provides the option to create cookies on the service IDs, but it still recommends everyone to use Cname, since ITP (Intelligent Tracking Prevention) discussed above, makes all these cookies very short lived.

# Seems like everything is working fine, or is it?

It does seem like that browsers have started identifying cookies being set through Cname implementation and are now acting against it. The brilliance of Cname was that it was fooling browsers, but Safari and Webkit browsers are becoming smarter to identify the cookies using Cname, therefore it is very device specific right now. Brave browser has already implemented blocking of Cname cookies, which means that it is able to identify alias domains, and match with it’s repository of domains, if it matches to any of the tracking website, the hit gets blocked, and hence the hit will never reach tools like Adobe Analytics.

# How does it identify Cname implementation?

The details are not yet widely available, but according to what I have been able to find, ITP tries to find out which trackers can have cross site tracking capabilities. Let me again give you an example. On sites abc.com & xyz.com, we have Adobe Analytics implemented. Which means their alias domains will look like abc.adobeanalytics.com and xyz.adobeanalytics.com. So, ITP will check the domain, not the subdomain, hence adobeanalytics.com. It will then see where trackers created on this domain, will be able to track user activities on more than one website, in this case, on both abc.com and xyz.com, if yes, then it will put the domain in cross site tracking list, and will flag this as Cname implementation.

Again, this is what I have been able to understand, but we can look into it little more to take a peek under the hood. We can debug these cookie logs, but for that we will have to have desktop/laptop version of Safari with us, because mobile version doesn’t let us debus the logs.

But essentially, I think this is how they are flagging all tracking website, through cross domain tracking capabilities, whether that be Adobe or Google Analytics.

# What about Facebook?

I am covering FB in a different section because they can’t be confused with cross domain tracking website. Now, when you go to facebook.com, it will set it’s own cookie, hence first party cookie. When you visit abc.com, and open an FB window to share something in an iframe, then fb.com through the iframe, reads the fb.com cookie, not abc.com cookie. This is how facebook till now is escaping the cross domain tracking feature, because essentially it’s not setting any cross domain cookies, it’s just reading cookies across domain, through iframe load.

This can again be a strategy for bypassing the current third party cookie issues, by ensuring that we set up third party cookies, and use logins and iframes to read these cookies. But we need to make sure that these iframes are intentional click iframes, else pop up blockers will stop them from getting loaded in the first place.

# What are data regulations?

There are 2 major regulations that we have to account for, and though there are other countries coming up with their own versions (including India), but essentially, pretty much every regulation has similar flavor to it.

* GDPR: General Data Protection Regulation agreed upon and ratified by European Parliament
* CCPA: California Consumer Privacy Act, for California consumers

These regulations basically guarantee following rights to consumers of these geographies:

* Right to be informed
* Right of access
* Right to rectification
* Right to deletion
* Right to restrict processing
* Right to data portability
* Right to object
* Rights around automated decision making
* Right to opt-out
* Right to non-discrimination

The most important part of these regulations is that most of the laws are territorial in nature, which means that every sovereign country can make laws affecting it’s people, which will be valid in the boundaries of that specific sovereign nation. In case of these data regulations though, these are extra territorial laws, which means that these regulations have to be followed by organizations outside the boundaries of the nations implementing these laws, and can also be enforced by various means.

Since the penalties of these laws can be really steep, so it’s imperative for us all to adhere to these regulations. For example, in case of GDPR, the fines can be in tune of 20 Million Euros or 4% of global revenue of the organization, whichever is higher.

# How would these regulations impact our data?

These regulations are designed to limit the collection, usage and processing of data, in a way that we can justify the users, and in case users do want to exercise any of their rights, we should be able to help them do the same. This means that we need to have a greater control over the data that is being collected, sources from where it is collected, data pipelines from where this data is flowing, centers where this data is processed, application of this data and all other touchpoints where data might be used in.

This kind of control is easier for us to create and monitor in case of first party data sources, but in case of third party data sources, it becomes much more difficult for us, since we rarely have a view of the way this data is collected or the algorithms used to process this data. For this reason, us, as data processors, third parties, as data collectors, and the platforms selling this data, as data sellers, have built in checks and balances for us to create controls. Cost of these checks and balances are accuracy and detail of data available to us. For example, now we might not be able to identify which car a specific user from Houston, Texas, might have, but we can still try to target the anonymous users from Houston, Texas, who have black Lexus.

# What can be our solutions in this current scenario?

One of the ways is to anonymize users, and ensure that all the processes, tools or technologies consuming data coming in our systems are created and provisioned to use anonymized data. Which means our marketing strategies have to be defined, even if in more details, on persona level, not on the level of individual users.

Also, we need to ensure that we increase our reliance on first party data, than on second or third party data. This can be done by creating our marketing strategies to entice users to visit our first party platforms, and once they enter our platforms, they should be able to login and share their information with us. Though login through FB, Google etc. is pretty useful, but it will be better if user registers through our own registration forms, so that we can collect more reliable information in our own systems, rather than relying upon the ‘walled gardens’ of Facebook, Google, Amazon or Apple.

# How server is the issue for us?

On the basis of information collected form the team, which is a very small sample, not really reflective of the entire Data CoE, it seems that currently we are actually using first party data in almost 80% of cases. Third party data usage is actually not as widespread in current processes, and even when it is used, it’s not always super critical.

# What’s next for us?

I think that there can be 3 high level solutions that we can build for our clients:

1. We help our clients create strategies to get users on first party platforms, and enable these platforms to collect data, in the right manner, after right consent, and store the same in meaningful structured formats
2. When users are coming to our platform, we do have a standardized structure of data in place, whether through data pipelines, meta data or query parameters, which will help us to join & merge data from these data sources. This merge will be more deterministic in nature.
3. We try to develop our probabilistic models to match records of users coming from different, disjointed sources
4. Own DMP/CDP set up is something that clients can look into, where primary IDs and first party cookies can be created, which can then be targeted by publishers.
5. Create guest and ‘member’ sections, so that people have incentive to give consent to put cookies on your website, thereby creating first party cookies.

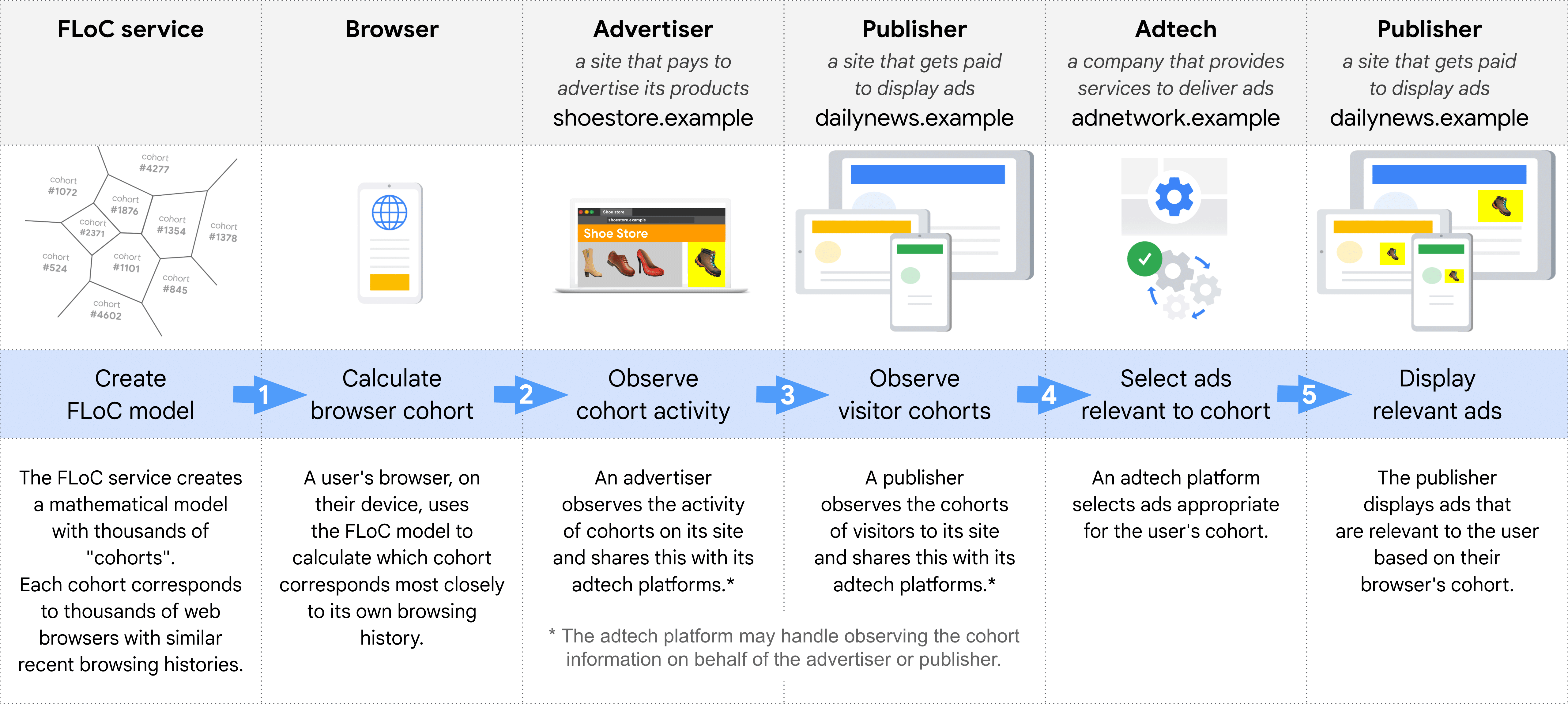
One of the examples to do this deterministic and probabilistic matching is a process of ‘Householding’. This essentially means to identify individual users belonging to one decision making household unit, with varying degree of confidence. This is very common method used in BFSi sector, since loan rates, interest on certificate deposits, can vary depending upon a person’s credit history etc. also, one household can have multiple services/products with a bank, for example, bigger banks aim for the metric of products per household, which can go from 2-4. So, there is a need to ensure that we reach to one household with consistent information.

# Google Privacy Sandbox

Since Google has decided to do away with 3rd party cookies, the idea can be to use browser to develop functionalities like SSO and personalized ads. For marketing and data teams, we will be concentrating on the following services:

## Federated user categorization & data analysis

FLoC: Enables ad selection without sharing individual browser details. As a user moves from one website to the other, FLoC algorithms will keep on calculating the interest of users, and as browsers work in the federated manner, thousands of browsers will categorize users on the basis of their interests. All calculations and identification of ‘interest cohorts’ will happen within browser, rather than centralized Google servers.



Learn more about federated learning: <https://www.tensorflow.org/federated/federated_learning>

## Understanding user interests from across site behavior

Fledge: API designed for on-device auction by browsers to choose relevant ads. Here user’s browser stores the advertiser defined ad groups and the auction is conducted locally on the browser depending upon business logic, buyer/seller data and user’s interest group.

An advertiser can’t learn about the pages a user visited or the identity of the user, ad selection is done on user’s browser only not on publisher or adtech’s platform.

# How can we use these templates to build our analytics?

We can take a couple of cues from the technology solutions discussed above to do the following:

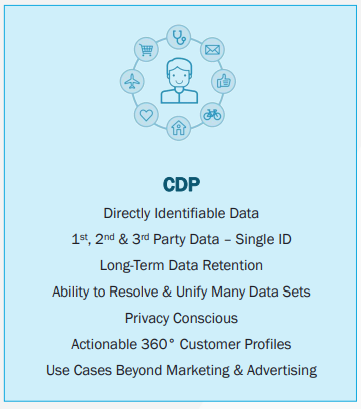
1. Use first party cookies and identifiers for our users
2. Provide benefits and functionalities to users for registration and login
3. Identify user interest groups through in browser federated learning algorithms, TensorFlow is the best option for us in the same.
4. Client’s interest group/user identification data can’t be merged with ad data, so will have to brainstorm on the same.

# Identity resolution in case of cookie blockers

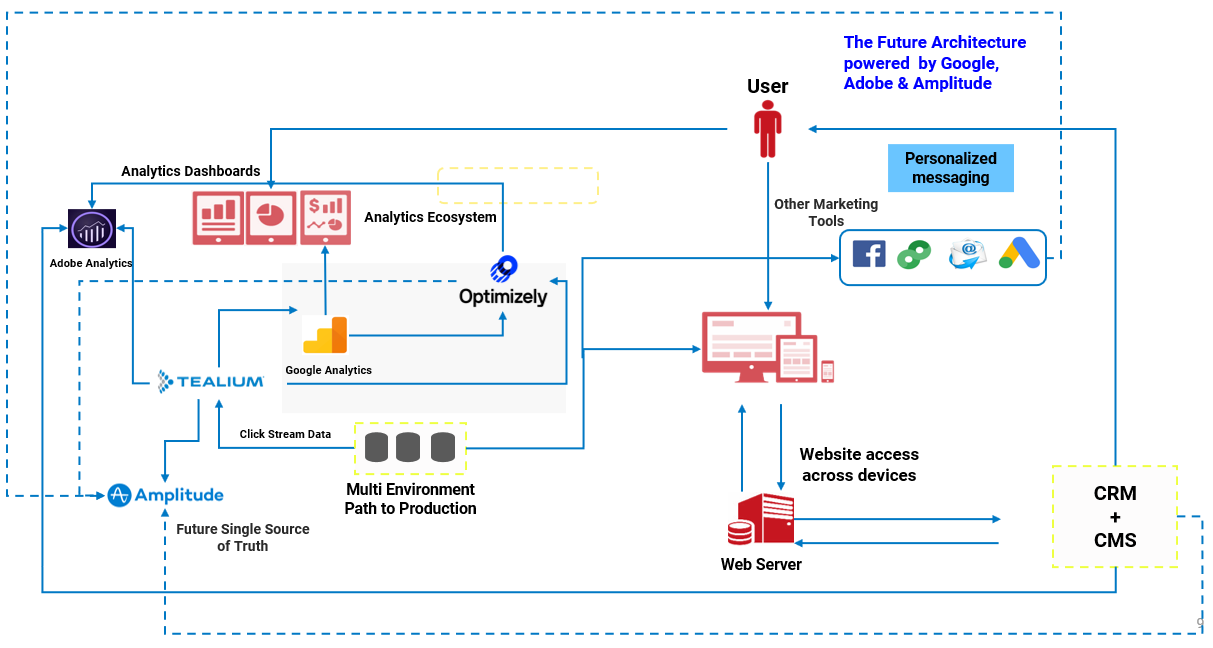
In case cookies are blocked, we have to identify used through the below methods:

1. Probabilistic: Through IP Address mapping
2. Deterministic:
   1. Email
   2. Walled gardens like FB, Google etc
   3. Logins
   4. Order ID
   5. Customer ID

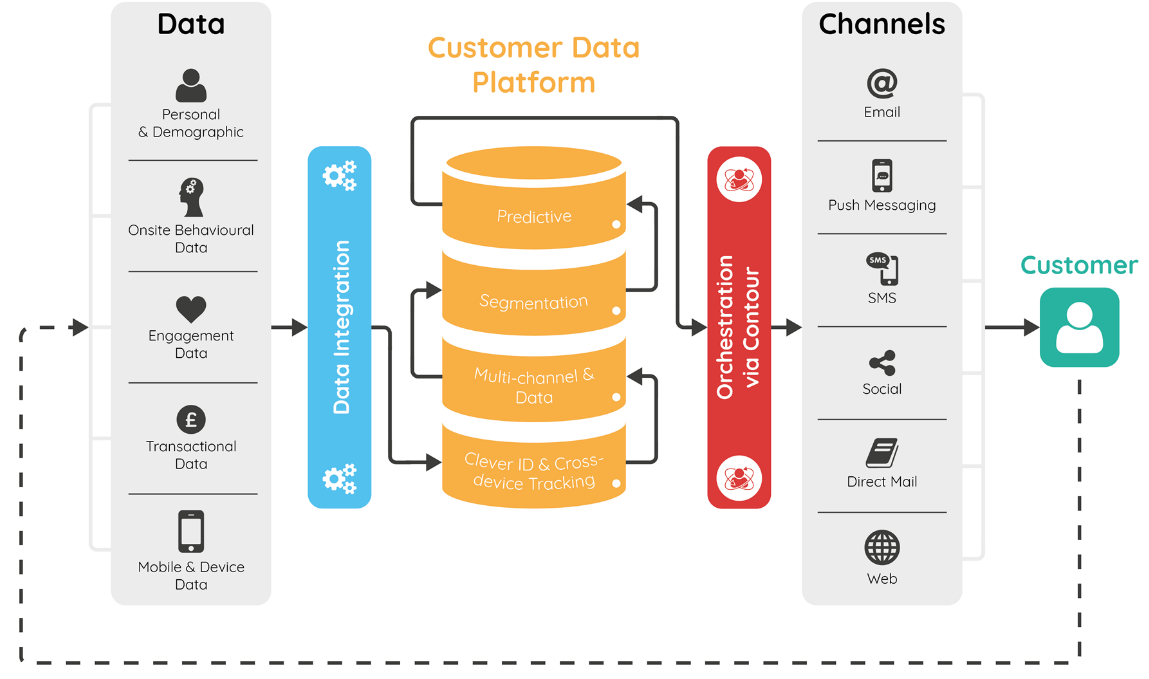
Implementing these methodologies in traditional data warehouses or databases might be highly cumbersome, Customer Data Platform (CDP) can be much better to implement user identification:



# Current & Future Analytics arch



In future we expect Amplitude to replace Adobe Analytics as single source of truth. We have to think about the prospect of implementing CDP within Amplitude



# Some other identity based tailored solutions

Google is not the only one who are working on identity resolution without 3rd party cookie data, there are a few more:

## Unified ID 2.0

Tradedesk media buying platform is platform is leading the initiatives with their open-source Unified ID 2.0.The way UID 2.0 works is - Once you have integrated the Tradedesk API with your website, after a user logs into the website their email id ,will be encrypted and will be shared with the Tradedesk platform. The regular rotation of decryption keys will provide extra security to user’s data. All of this, will make it impossible for anyone to reverse engineer the hashed keys back to email addresses. These hashed and salted identifies can be used by the Tradedesk’s partners to understand user’s journey better.

## SPARROW (Secure Private Advertising Remotely Run On Webserver)

Criterio’s framework proposal for replacement of third party cookie retargeting. This framework will still involve outside servers, an independent third party, a Gatekeeper. This Gatekeeper will act as an independent third party that will oversee the processes between publishers and the advertisers. It will be able to give better control of data to the users and still enable advertisers to retarget users.

## Fatbrick ID

Neustar has come up with a deterministic identifier (Fabrick ID) , which is an identity based customer cloud solution to provide a secure and future proof connection between advertisers and audiences. Being a pseudonymized user token (Fabrick ID), it will help brands to create cross media linkage without affecting user’s privacy.

# What is the solution?

We don’t have a readymade solution available in the market which can solve the issue, and probably due to the same reason Google has also decided to postpone the decision to discontinue 3rd party cookies by a period of 2 years [source](https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones). At this point of time developers, marketers and analysts have to come together to test out various APIs provided by Google and other participants of the initiatives, to prepare our marketing teams for the same.

You can read more about the Google’s Privacy Sandbox from [here](https://www.privacysandbox.com/), and test out the APIs by yourself to see how the ecosystem will come together for the marketers.

# Frequently Asked Questions

**What is happening?**

Google has decided that it will block all third party cookies on its Chrome browser in 2023. We don’t have exact dates, but considering around 60%-70% of traffic on any digital platform comes through Chrome, and the dependence on Google and its ad services in the space of digital marketing, makes it imperative for all advertisers to take it very seriously.

**What does it mean? Why should we lose sleep over it?**

Cookies are nothing but small scripts which retain a certain value for a specific period of time. In the world of digital marketing it stores identifiers like client ID, visitor ID, which can then be persisted to record and analyze user behavior on the platform. These IDs are also used to marketing by the concept of cookie synching, where you can send your audience lists to the ad networks, to be targeted on various publishers’ website. All this works on the very foundation of third party cookies. This change implies that there will be an impact on both, marketing, as well as measuring the impact of the marketing on digital platforms, leaving advertisers in the flux.

**What sort of impact can we expect on us?**

Impact of such a change will be felt on all stakeholders involved in the process of digital marketing:

1. *Advertisers*: They will lose the power to merge their data with ad networks and Publishers’ data for remarketing and targeting purposes, since advertiser’s data will now be third party data for publishers and ad networks. Advertisers will also have to build on differential privacy, and 1st party data sets, through consent of users, to better create their marketing strategies, since audience attribute enrichment process through 3rd party datasets is bound to take a hit.
2. *Publishers*: They have to not just use their own 1st party data for better targeting and better results for the advertisers, but also will have to come up with AI/ML techniques to categorize users in multiple intent based cohorts to fill the gap from 3rd party cookie data
3. *Marketing Agencies*: Not only marketing agencies will be impacted, but now are duty bound to enhance their roles from just execution/operations to partnership in these uncertain times. These agencies will have to ensure that websites are tagged properly with tags like Universal Event Tracking (UET) of MS, or Gtag of Google, to ensure that they are getting their own 1st party information.

At the same time build algorithms to cluster user groups in various intent based cohorts to further enhance the value of data collected within these tags and services. At the same time the need of identity resolution graphs with the help of deterministic and probabilistic matching, will take a centerstage especially while measuring the impact of marketing efforts in absence of third party cookies.

1. *Ad Networks*: They will have to ensure that they are also able to collect information of users coming on different platforms, categories them on their behavior and intent in the best manner possible, because this will become the very backbone data for targeting users in absence of 3rd party cookies.
2. *Measurement*: Whereas everyone is talking about impact on advertisers and publishers, very few are talking about its impact on measurement of marketing strategies in traditional reporting environment. Considering we are most likely to lose the identity of users, hence attribution will severely be impacted, especially multi touch attribution algorithms, whether out of box or custom algorithms. This means an extra layer of identity resolution has to be first applied to accurately calculate the attribution of conversions, with the help of vendors like Tradedesk, Liveramp etc. or by our own custom identity graphs, create don the back of 1st party data in Customer Data Platforms (CDPs).

All this means that measurement of marketing efforts will now happen on modeled datasets, rather than lift and shirt of ad data from any publisher/platform into Excel/BI reports.

1. *Tools and technologies*: The biggest impact will be felt on DMPs (Data Management Platforms), used for curating audiences which could be targeted on various destination. Since this system worked on the foundation of cookie synching, with third party cookies getting deprecated, will severely hamper their capabilities, unless they come up with better and more evolved way of identity resolution.

CDPs can become extremely integral part of the digital marketing ecosystem, but only if they are equipped with identity resolution algorithms. Customer identification and householding with the help of deterministic and probabilistic algorithms will become a key part in devising the marketing strategies. In addition to this there is a need to cluster users on the basis of their intent for contextual targeting. This is again something that can be built over CDPs, to further enhance their value in the ecosystem.

Ad Networks like Microsoft are going to use hashed email address, addressing privacy and still providing better targeting capabilities, for identity resolution.

Tradedesk uses a Universal ID (UID) to identify users and retains the capability to target them in post 3rd party cookie world.

Liveramp has the capability to match PIIs into RampIDs for targeting and can also provide measurement information mapped with AbiliTec ID

**If third party cookies are deprecated, what could be new data sources?**

Marketing most likely will now depend more upon contextual messaging, than any new third party data source. This means that reliance on 1st party data sources is now much more than before. Earlier data used to be a by product of a transaction, like entering email address/creating account while buying. Now data collection has to become a transaction/ one of the primary goals itself. Incentivizing users to share their data, use their loyalty cards has to be become a part of the culture for advertisers. Showcasing personalized offers, rather than same generic offers running on banners, thereby enhancing user experience, and giving a reason to users for sharing their information has to become the key.

For the same reason, second best thing can be someone else’s first party data, that is second party data. In this data coming from different retailers, in a clean room kind of a set up can be another vital data source to further enhance own data sets of advertisers.

In addition of data sources, application of AI/M: algorithms to create audience cohorts, not just on acquisition dates, but also on intent, can further improve both marketing strategies and measurement of these strategies.

**How Microsoft is helping address this challenge?**

Microsoft has recommended use of UET, Universal Event Tracking, to circumvent the possible challenges. This is a simple tag that can be created within Bing Ads platform, and can be added across your website. This tag will help MS collect information about users on its own, thereby creating first party data for better targeting. You can also create 5 kinds of goals on Bing Ads for conversion tracking through UET, like:

1. Destination goals
2. Duration goals
3. Pageviews per visit goals
4. Event goals
5. Mobile App Installation goals

In addition to UET, Microsoft also uses hashed email address to bolster its targeting algorithms.

**How Google is going to address challenge created by itself?**

Google is just going by privacy concerns of users and regulations by institutions across the globe. For the same reason it has been gearing towards the solution as well for quite some time now. Google’s own Gtag when implemented on the website collected information about users, and thereby enriching its own first party data.

Google also is working on 2 algorithms, FLoC (Federated Learning of Cohorts )and Fledge. FLoC is geared more towards interest and behavior of users whereas Fledge deals with building audiences for advertisers.

Google has categorically confirmed that they will NOT provide any identity resolution like PII graphs based on people’s email addresses, because they don’t think it meets consumer’s expectations from the evolving systems.

**What should I do right now?**

For one, the first thing we recommend is what you should NOT do, that is, don’t panic. Cookies were getting deprecated all this while, Safari and Mozilla Firefox have been doing it for a while now, so we do have systems in place. Of Course, Chrome, considering the size and volumes of traffic it generates, will have a much bigger impact, but we still do have systems in place. There will be initial issues, but we all are in this together, and the existing systems will also evolve and develop to counter the challenge of addressability.

Harvesting 1st party data sets has to be the step 1 of gearing yourself to address the upcoming issues. Partnering with right agencies and vendors can be close second. Agencies and teams having well entrenched processes and tools to create identity graphs, intent based cohorts, with deeper understanding of the evolving space will help a lot to not just address the upcoming challenges, but also outperform your previous set benchmarks.

Business with longer purchase cycles are expected to feel the impact more than the ones with shorter purchase cycles, just because of the reason of persistence of information, but again, as time goes by, as systems evolve, this should get better.

Current anxiety is due to the nature of change, than the change itself, like the number of unknowns, the number of players, and a general ambiguity of how they might come together, are the reasons. For us at Publicis Sapient, cookiepocalypse is a definitely an exaggeration of a challenge that is sure big but not unsurmountable.