

Tracking Anonymous Customer Journeys

C Bushell 2025

Understanding and tagging the full customer journey even when purchases happen long after an initial touch – requires a strategic blend of technology and data management. Modern platforms like Google Analytics 4 (GA4), Meta’s Business Suite (Facebook/Instagram), Salesforce CRM/Marketing Cloud, and Google BigQuery each offer tools to persistently identify users, log events across channels, resolve identities across devices, and attribute conversions back to early marketing touchpoints. Below, we break down how each platform supports long term customer journey tracking, with real world techniques from brands like Starbucks, Gymshark, and Netflix. We then highlight advanced methods (probabilistic modeling, server side tracking) and conclude with practical takeaways for implementing a long term attribution framework.

Introduction: The Challenge of Long Latency Conversions

Most analytics and ad platforms focus on short attribution windows (e.g. 7–30 days) for linking ads or clicks to conversions. In reality, consumer behavior is nonlinear and can span months or years. A customer might see a social ad today and not purchase until much later, making the conversion seem “random” unless early touchpoints are tracked and tagged properly[1].

Traditional last click models fail to credit these early brand impressions. To bridge this gap, marketers are shifting to multi touch attribution models and longer lookback analyses – in fact, 41% of organisations now use attribution modeling to measure ROI[2]. Achieving this at scale demands persistent user identifiers, comprehensive event logging, cross channel identity resolution, and often retroactive analysis in a data warehouse. Crucially, any long term tracking strategy must respect user privacy and rely on consent based, first party data[3].

Google Analytics 4 (GA4): User Centric Event Tracking and Attribution

GA4 is built around an *event based* measurement model and offers improved capabilities to stitch together user journeys across devices and sessions. Key features include:

Persistent User Identification: GA4 allows using a User ID (set by the business for logged in users) to track individuals across devices[4]. If a known user is logged into your site/app, all their events can be tied to this ID. GA4 also assigns an anonymous device ID (`client_id`) for each device; by default this only ties behavior on that single device[4]. However, GA4’s *blended identity* setting will automatically stitch sessions across identifiers – it first checks for a User ID, if not available falls back to device ID, and if neither is present (e.g. cookies declined) it uses modeled data[5][6]. This modeling uses the behavior of similar consenting users to fill gaps for users who can’t be identified, a probabilistic method to avoid losing the entire journey[6].

Event Logging & Early Touch Tagging: Every interaction can be logged as an event in GA4, enabling granular tagging of each touchpoint. Marketers should configure events for key funnel steps (page views, product views, add to cart, etc.) and capture campaign parameters (UTM tags, referral source) on entry events. These events are time stamped and tied to user identifiers, so even if a purchase occurs much later, the prior events remain in the user’s history. GA4’s data retention for user level data can be extended (up to 14 months by default in the UI, or longer via BigQuery export), which helps in analysing long latency conversions. For example, Gymshark (a fast growing fitness retailer) overhauled its analytics with GA4 to focus on *user level journeys*

instead of one off sessions. They recognised that purchase paths were “increasingly fragmented and nuanced” and thus prioritised first party user data and cross platform tracking (web + app) to get a holistic view[7]. Gymshark even changed its success metrics from session based conversions to user conversion rates, reflecting a shift to long term, cross session engagement[7]. This GA4 implementation, aided by a first party login system, allowed Gymshark to identify repeat interactions leading up to a sale and optimise accordingly. As the team noted, transitioning from session based to user based analytics provided “a more fluid and holistic way to measure customer value and cross channel performance”[8].

Cross Device and Cross Platform Stitching: With User ID enabled, GA4 will unify a user’s interactions across multiple devices into a single journey. For instance, if a customer browses on mobile but purchases on desktop (logged in on both), GA4 attributes the desktop purchase to the original mobile touchpoint in its user centric reports[9][10]. GA4 also utilises Google signals (if enabled) to help recognise users across devices when they are signed in to Google, supplementing first party IDs. In practice, this means early upper funnel events (like viewing a product on the app) can be tied to later lower funnel events (like purchasing on the website) for the same user journey. This cross device unification relieves marketers from guessing whether a late conversion on one platform was influenced by a prior touch on another[9].

Attribution and Retroactive Credit: GA4 uses a data driven attribution model by default for conversion reporting. This model algorithmically assigns credit to multiple touchpoints (paid and organic) in a user’s path, rather than just last click. As long as the conversion happens within the attribution window (typically up to 90 days for many channels), GA4 will credit earlier steps (e.g. an ad click or email open that happened weeks before the purchase) based on their influence[11]. Marketers can also analyse conversion paths in GA4’s Advertising reports to see the sequence of touches that lead to conversion, even if they span multiple sessions. However, if a purchase occurs outside the standard lookback window or without a persistent ID connection, GA4 alone may not link it to the original touch. For truly *long term* windows (e.g. a conversion 6+ months later), exporting GA4 data to a warehouse is advisable (see BigQuery below). GA4’s BigQuery integration exports raw event data (nearly real time), which data teams can query to perform retroactive attribution over any time span[12]. *Note:* By default, GA4’s BigQuery export is keyed by device (each record has a device based user_pseudo_id), so stitching the same user’s data across sessions/devices requires querying on your own identifiers (e.g. linking records by User ID)[13]. This flexibility lets analysts reconstruct a user’s entire history (e.g. all impressions and visits in the past year) once they become a customer.

In practice: GA4’s ability to combine first party identity data, comprehensive event tracking, and flexible attribution windows makes it a foundation for long term journey tagging. Gymshark’s adoption of GA4, for example, enabled them to “[see] the purchase journey from end to end” and discover pain points and influences across many touchpoints[14][15]. The brand found that user behavior differed vastly between channels (web visitors might come 2 3 times a month, whereas logged in app users engaged 5 10 times a week), insights only visible through unified cross channel data[16]. Using GA4 Explorations and BigQuery, they identified counterintuitive trends – for instance, customers with too many items in their wishlist were less likely to convert, leading Gymshark to adjust the UX and improve conversion rates[17][18]. These kinds of findings illustrate the power of persistent journey tracking: by logging every interaction and

tying it to users over the long term, brands can attribute outcomes to early behaviors and optimise the full funnel, not just the last step.

Meta Business Suite (Facebook/Instagram): People Based Attribution Across Devices

Meta's advertising ecosystem (Facebook and Instagram, managed via Business Suite/Ads Manager) is inherently people based, meaning it relies on user login identity to track ad exposures and actions across devices. This gives Meta a strong ability to tag and attribute long customer journeys, as described below:

Persistent Identity via Login and Advanced Matching: Facebook has billions of users who stay logged in on multiple devices. When someone interacts with an ad on one device and converts on another, Meta often can match that activity to the same person so long as one of their devices or input data is recognised[19]. As Meta's documentation notes, *"people are often logged into their Meta accounts from multiple devices"*, which allows linking of a click on a mobile ad to a later desktop website purchase[20]. In addition, Meta provides Advanced Matching for the Meta Pixel and Conversions API. This feature lets advertisers send personally identifying information (PII) – e.g. a hashed email, phone number, or name – along with conversion events. If a user purchases on your website and you collect their email, the Pixel/Conversions API can hash and send that email to Facebook, which attempts to match it to a Facebook user profile. This deterministic match can bridge gaps when browser tracking isn't reliable. For instance, if a user saw an Instagram ad, didn't click, but later visited your site directly and signed up with the same email, advanced matching can connect that conversion back to the ad impression. Meta's platform uses these data to resolve identities across channels – effectively creating a unified view of the customer by leveraging its huge logged in user graph[21][22]. (According to one analysis, Meta's advanced matching and login data make cross device attribution "very possible" even when third party cookies or URL parameters fail[20].)

Event Logging via Pixel and Conversions API: To tag customer journey events for Meta, businesses implement the Meta Pixel on websites (and SDK in apps) to log standard events (PageView, ViewContent, AddToCart, Purchase, etc.) as well as custom events. Every time a user visits or does something on your site, the Pixel can fire an event that is attributed to that user's Facebook profile if possible. These events include a unique user ID (if available) or Facebook cookie ID to tie back to ad exposures. For more robust tracking (especially post iOS14), Meta strongly encourages using the Conversions API (CAPI) – a server side method to send the same events from your server to Facebook. Server side events are not subject to browser blocking and can contain richer customer data (like order values or user IDs) in a secure way. By combining client side Pixel data and server side CAPI data, Meta aims to capture as many journey touchpoints as possible, even if the user has tracking blocked on the browser. For example, Starbucks leverages its loyalty and mobile app data to feed Facebook with custom audiences and conversion events: when a loyalty member redeems an offer in store, Starbucks can upload that as an offline conversion tied to that member's email/ID, which Facebook might match to a prior ad the user saw. This way, a purchase that occurs in a different channel (brick and mortar store) days or weeks after an online ad impression can still be attributed back to the ad via persistent identity. (Starbucks' broader cross channel strategy – integrating mobile app, email, and in store via the Starbucks Rewards ID – exemplifies how a brand can maintain continuity in user identification. With over 31 million active loyalty members, Starbucks uses that unified ID to coordinate promotions and track engagement across touchpoints[23]. The

payoff is clear: by seamlessly linking channels and data, Starbucks saw a *40% boost* in loyalty membership from 2018 to 2022[24], indicating that customers respond to the cohesive experience and the company can better measure their behavior over time.)

Attribution Windows and Cross Channel Credit: Meta Ads are optimised and reported using configurable attribution windows. The default “Standard” setting is a 7 day click and 1 day view window[25]. This means if a person clicks an ad and converts within 7 days, or views (but doesn’t click) an ad and converts within 1 day, Meta will attribute that conversion to the ad by default. Historically, Facebook offered up to a 28 day post click window, which was removed as default after iOS 14 but is still available for manual reporting comparisons[26]. Advertisers can compare 1 day vs 7 day vs 28 day attribution results to understand how many sales come from longer consideration cycles[11]. Importantly, these windows *don’t require continuous cookies on the user’s browser*, because Facebook ties conversions to users/server events. If the same person (as recognised by Facebook) converts even weeks after an impression, and it falls in the chosen window, the credit is given. Meta’s cross device tracking means the “conversion device” can differ from the “impression device” – reports explicitly show which device saw the ad vs. which device was used to convert[27]. Behind the scenes, Meta unifies those events by user. This is critical for tagging multi channel journeys; for example, you might see that many mobile ad impressions eventually led to desktop purchases – insight you’d miss without cross device attribution.

Probabilistic Modeling for Missing Data: In the post ATT (App Tracking Transparency) era, a significant portion of user activity (especially on iOS devices) is not directly trackable. Meta has responded by deploying conversion modeling to fill the gaps[28]. Conversion modeling uses machine learning on aggregated data to estimate conversions that cannot be observed due to privacy restrictions[28]. For instance, if 100 people clicked a Facebook ad on their iPhones and later many made purchases but didn’t allow tracking, Meta will statistically estimate how many of those purchases to attribute to the ad based on patterns from those who are trackable. These *modeled conversions* are then included in the 7 day/1 day attribution results to provide a fuller picture[28][29]. In mid 2021, Facebook introduced updates to model even the 7 day window for opted out users (previously, opt outs were only counted in a 1 day click window)[30]. The effect is that Meta’s reporting now adds estimated conversions into its 7 day click figures, “helping marketers account for lost data” from devices where direct tracking is missing[29]. This probabilistic approach ensures that long latency conversions aren’t systematically under counted just because the trail went cold after the initial click. The accuracy is continually refined by comparing modeled outcomes to “ground truth” conversions Meta can observe[31]. While not perfect, it gives advertisers confidence that even if a user’s journey is fragmented or partially invisible, the attribution model will allocate some credit to earlier Facebook touches with a measurable degree of certainty[28].

In practice: A brand running multi channel campaigns should feed all available conversion events into Meta (via Pixel/CAPI, including offline sales uploads) and leverage Meta’s identity matching. For example, an ecommerce brand like Gymshark (known for heavy social media marketing) could use the Conversions API to send purchase events from its backend, including customer email or phone. If a Gymshark customer buys a product a month after seeing a Facebook video ad, Gymshark’s server will send that order with the email; Facebook can hash it, recognise the user, and attribute the sale to the video ad (as long as it’s within the 28 day

reporting window). Many D2C brands operate this way, effectively building their own CRM to Facebook feedback loop. Meta's own documentation highlights that combining data sources improves attribution: "advertisers can provide first party data via the API" (like emails) to bolster cross device matching[21]. The key is persistent tagging of each user and event – by using a login or unique identifier and passing it along – so that even delayed purchases have a breadcrumb trail back to the initial ad impression. Meta's platform handles much of the heavy lifting in matching and modeling, giving brands a relatively straightforward way to measure long term ad impact (within a month or so) and across disparate devices.

Salesforce: CRM Centric Journey Tracking and Multi Touch Attribution

Salesforce, encompassing its CRM (Sales Cloud) and Marketing Cloud (including tools like Journey Builder, Pardot, and Marketing Cloud Intelligence), approaches customer journey tagging from the angle of a unified customer profile. The emphasis is on capturing every interaction (marketing, sales, service) tied to a known individual and attributing downstream revenue to those interactions. Here's how Salesforce supports long horizon journey tracking:

Persistent Customer Identity in CRM: Salesforce's core strength is as a customer system of record. Each customer or prospect typically has a unique ID (Lead or Contact ID in Sales Cloud) which all interactions can be linked to. When a person fills out a form, joins a loyalty program, or otherwise identifies themselves, Salesforce records them as a contact and can store their email, phone, and even anonymous web cookies (via Marketing Cloud/Pardot tracking) associated with that profile. This means even if a user starts out anonymous (browsing a site), the moment they provide an identifier, you can retroactively link their past behavior to their new profile. Many Salesforce marketing implementations use this strategy: an anonymous visitor ID (from a cookie or Pardot tracker) is tracked over time; when that visitor eventually registers or makes a purchase, all their previous pageviews or email clicks (that were tagged with that visitor ID) get merged into the now known contact's history. This persistent identity is often maintained for years in the CRM. For example, Starbucks uses a CRM centric approach via its Rewards program – each member's interactions in app, via email, or in store are tied to their loyalty ID (which is essentially a CRM contact). That allows Starbucks to see that you might have first signed up during a promotion last year, engaged with emails in the spring, and only this fall converted by buying a new product. All those touchpoints are unified under one profile, giving a longitudinal view of the customer. (This aligns with what Malhotra et al. call "*Social CRM*" – integrating customer data across channels over time[32].)

Comprehensive Event Logging (Marketing + Sales Touchpoints): Within Salesforce, every marketing campaign or touchpoint can be logged as a record attached to the customer. Salesforce has a "Campaign" object to represent marketing efforts (an email blast, an ad campaign, a trade show, etc.), and when a customer interacts, they can be marked as a Campaign Member. For digital interactions, Salesforce Marketing Cloud and Pardot provide tracking tools: for instance, Pardot uses tracking links and cookies to record email click throughs and page visits, while Marketing Cloud's Journey Builder can listen for events like email open, SMS click, or ad impressions (via Advertising Studio). Additionally, offline interactions like sales calls, store visits, or support tickets get recorded in the CRM. All of these become data points on the customer timeline. By tagging each interaction in Salesforce, a company can later query "Which touches did this customer have before purchase?" even if those touches occurred many months or years prior. This is crucial for retroactive attribution: if a sale closes today, Salesforce's data

might show that the customer first clicked a Facebook Lead Ad 12 months ago, then attended a webinar, then had an outbound call with a sales rep. In other systems those events might live separately, but in Salesforce they roll up to one profile. For instance, if Netflix were using Salesforce to track marketing (hypothetically), they could log that a person watched a free content teaser (campaign A), then later received a sign up email (campaign B), and only subscribed after seeing a particular show launch. Having those events in one place allows attribution of that subscription not just to the last touch, but to the earlier influential touches as well.

Cross Channel Identity Resolution: Salesforce recognises that customers interact via multiple channels (email, web, mobile, ads, in person) and often switch devices. To address this, Salesforce's platform can use both deterministic and probabilistic matching techniques[33]. Deterministic matching is straightforward in Salesforce: if the same email or phone is used, it's the same person. Tools like Salesforce CDP (Customer Data Platform) extend this by ingesting data from various sources (e.g. e commerce transactions, mobile app usage, ad impressions) and matching records through common identifiers (account logins, device IDs when available, etc.) to build a unified profile. Probabilistic matching might be used in cases where a direct identifier isn't present – for example, matching an anonymous website cookie to a known user by comparing behavior or IP/location data when a login eventually happens. Salesforce's documentation suggests using “cross device tracking solutions” and both probabilistic or deterministic methods to connect interactions that span devices[33]. In practical terms, many companies integrate Salesforce with their web analytics or ad platforms: using lead tracking pixels or campaign parameters to tie an ad click (with a unique ID in URL) to a Salesforce record once they convert. Encouraging users to identify themselves (e.g. sign up for a newsletter or loyalty program) early in the journey is a common strategy – it converts unknown visitors into known contacts, after which Salesforce can unify all future and past touches under that contact. Starbucks exemplifies this by heavily promoting account sign ups (loyalty sign in) for WiFi or app use; once signed in, every action is captured against their profile, whether it's an in app coupon view or an in store purchase via the app.

Multi Touch Attribution and Campaign Influence: Salesforce provides out of the box tools for attributing revenue to multiple touchpoints. The Campaign Influence feature in Sales Cloud allows you to associate an Opportunity (sale) with all the campaigns that a customer interacted with before the sale. Salesforce can apply different attribution models to these influences – for example, a First Touch model (credit goes to the first campaign that touched the customer), Last Touch, or Even Distribution across all touches[34]. If you have Salesforce Pardot (Marketing Cloud Account Engagement), there are configurable multi touch attribution dashboards that use standard models (first, last, linear) or even customisable weightings[35][36]. This means if a B2B lead was generated by a whitepaper download (Campaign A), then later nurtured via an email series (Campaign B) and finally closed by a sales call (Campaign C), you can set up Salesforce to assign, say, 50% credit to A, 20% to B, 30% to C (or whichever scheme fits your strategy). These attributions can be reported in Salesforce, tying actual revenue back to each marketing touch. The advantage is that it works on long sales cycles; a campaign can influence a deal many quarters before it closes and still get partial credit. Moreover, Salesforce encourages including offline touches in this picture – e.g. a trade show meeting or a direct mail response can be a campaign touch that is logged and credited[37]. By standardising how all channels' data feed into Salesforce, marketers create a central timeline for each customer.

Advanced Analytics and AI: For very complex journeys and large datasets, Salesforce offers tools like Marketing Cloud Intelligence (formerly Datorama) and Tableau/CRM Analytics (Einstein). These can ingest data from Salesforce and external sources to do advanced attribution modeling and even media mix modeling. Salesforce's guidance suggests using machine learning and predictive modeling to uncover patterns that traditional analysis might miss[38]. For example, an algorithmic attribution model (akin to a Shapley or Markov model) could be implemented to analyse which sequence of touches has the highest probability of conversion. Salesforce doesn't provide a built in algorithmic MTA in core CRM, but using Einstein Analytics or external tools, companies can perform such modeling on the Salesforce data. The goal is to quantify the impact of earlier funnel activities on eventual conversion, even if there is a long time gap. Salesforce's State of Marketing report notes that marketers are indeed moving this direction, investing in data integration and advanced analytics to achieve accurate multi touch attribution[2][39]. One recommendation is to continuously refine your attribution approach by feeding back results and running controlled experiments[38][40] – Salesforce's platform, with its unified data, makes experimentation (like A/B testing different journeys or comparing cohorts) feasible over long periods.

In practice: Companies using Salesforce should ensure that every digital campaign is connected to Salesforce either via tracking links or API integration. For instance, if Gymshark were using Salesforce (hypothetically), they might capture every ad click with a unique parameter that gets stored when the user signs up or purchases. Salesforce would then show that, say, a customer first came through *Campaign "Holiday Promo 2024"*, then interacted with *Email Series X*, and eventually converted in *March 2025*. All these would be logged under that customer. The attribution framework could then assign credit to both the Holiday Promo and Email Series for that sale. Another real world example is how many retail banks use Salesforce: a prospect might see a Facebook ad, then a Google search ad, then walk into a branch to open an account. By integrating ad data and in person CRM data, the bank can attribute that account opening to the earlier digital touches (perhaps using a unique offer code to tie the branch visit back to the online campaign). This aligns with Salesforce's advice to *incorporate offline touchpoints* and use both deterministic links (offer codes, emails) and probabilistic matching to connect the dots[33][37]. The upshot is a 360° customer view where even delayed conversions in a sales pipeline can be traced back to the marketing seeds that planted them.

BigQuery & Data Warehousing: Unified Data for Retroactive Journey Analysis

While platforms like GA4, Meta, and Salesforce each have their own tracking and attribution capabilities, a cloud data warehouse (such as Google BigQuery, Amazon Redshift, Snowflake, etc.) is often the linchpin of *long term, cross platform attribution*. BigQuery in particular is used by many brands in conjunction with GA4 and Google's marketing stack to store and analyse customer journey data without retention limits. Here's how BigQuery (and warehousing in general) supports persistent journey tagging and delayed purchase attribution:

Centralising Multi Source Data: A data warehouse allows you to merge data from all channels and platforms using your own business logic. For example, you can take raw web analytics events (from GA4's export), mobile app events, CRM customer records, email marketing logs, ad impressions, and offline sales, and put them into BigQuery tables. Using unique keys (such as a user ID, email, or even fussy matching on names), you can join these datasets to create a comprehensive timeline per user or per purchase. This is essential for tracking fragmented

journeys. One customer might appear in five different systems: BigQuery lets you link them into one record. Gymshark, for instance, exports its GA4 data to BigQuery[41] (as many GA4 users do) and could combine it with its e-commerce database. This would enable Gymshark's analysts to trace a single user from an ad impression in Facebook, to a website browse captured by GA4, to an email click recorded in their ESP, all the way to an order in their backend system. Without BigQuery, those events would live in silos. By warehousing them, Gymshark gains the ability to run SQL queries or machine learning to find patterns like "X% of users who eventually purchase had their first touch on our mobile app and not the website" – insights that drive long term marketing strategy.

Infinite Lookback Window & Data Retention: Unlike many analytics tools that might sample data or drop user level data after a certain period, a warehouse can retain years of history at the individual event level. BigQuery becomes the system of record for all impression and conversion data. This means if a user converts today, you can query all events from, say, the past 5 years to see if that user had any prior interactions. Indeed, researchers advocate this approach: only with a data warehouse can brands truly "track impressions through the full funnel, even over multi year timeframes," by retaining all those touchpoint records[1][12]. Netflix is a great example of leveraging big data for attribution. Netflix collects massive amounts of user engagement data on its platform and marketing channels, storing it in their data infrastructure (they use AWS/S3 and Spark, akin to a BigQuery approach). With this data, Netflix can perform sophisticated analyses like estimating how a given marketing touch (e.g. seeing a trailer for a new show) contributes to a user's decision to subscribe months later. In fact, Netflix published a research paper outlining a scalable methodology to estimate the incremental acquisition impact of content launches on subscriber sign ups[42] – essentially a form of attribution for a content "touchpoint" that might drive new customers over a long horizon. Achieving this required integrating content marketing data with subscription data and applying statistical models, something only feasible with a robust data pipeline and warehouse. The key takeaway is that with complete data retention, you can always go back and retroactively analyse the influence of earlier events on later outcomes. If your CMO asks, "What early touch in 2022 most predicted purchases in 2024?", having the raw data in BigQuery makes it possible to query and find an answer.

Cross Channel Identity Resolution (Custom): In a warehouse, you are not limited to the identity matching rules of any single platform – you can implement your own identity resolution schema. This might involve creating a unified user ID that links various identifiers: for example, connecting an email address (from CRM) to a GA4 client_id (from web cookies) when a user logs in, or linking a device ID to an email from a mobile app login. BigQuery can store mapping tables for these identities. Many brands set up an identity graph in their warehouse: a table that maps all known IDs and devices belonging to the same person. This can incorporate deterministic matches (exact email or userID matches) and even probabilistic matches (for instance, using machine learning to cluster devices that often behave similarly and might be the same user household). The end result is the ability to query the data by a stable customer identifier and get all events across all platforms for that customer. If Starbucks were to use BigQuery, they could import loyalty member IDs, website visitor cookies, and mobile device IDs and reconcile them. Then a single SQL query could pull every touch – from seeing a Facebook ad (matched via email in a custom audience) to an app open to a POS transaction – for a given customer over years. This level of unified tagging is the *holy grail* for long term journey analysis, and it typically lives in the data warehouse.

Custom Attribution Modeling and AI: With all data in one place, analysts can build bespoke attribution models that suit their business. BigQuery's power (and similar tools) allows processing large sequences of events to attribute credit. Two common approaches are rule based multi touch models (which you can implement via SQL – e.g. assigning 40% credit to first touch, 40% to last, 20% evenly to middle touches for each conversion) and algorithmic models. Algorithmic multi touch attribution often uses techniques like Markov chains or Shapley values to determine the importance of each touchpoint by simulating the removal of that touch or by learning transition probabilities in the customer journey. These calculations are computationally intense, but BigQuery can handle them on big data sets. Additionally, machine learning models (TensorFlow, XGBoost, etc., which can be run on BigQuery via BigQuery ML or external notebooks) can be trained to predict conversion based on sequences of touches. This is where probabilistic attribution comes into play: for instance, if you want to attribute value to an *untrackable brand impression* (like someone seeing a billboard or a TV ad), you might use a media mix model (MMM) or uplift modeling in the data warehouse. MMM is a regression based approach that estimates the contribution of each marketing channel to sales over time, independent of user tracking. It's essentially a probabilistic model at the aggregate level. Many advanced brands (perhaps Netflix included) supplement user level tagging with MMM in their warehouse to capture long term, diffuse effects of marketing that aren't directly linkable. BigQuery can house the data needed for MMM (e.g. spend, impressions, sales over time) and even run the regression models. Moreover, with the rise of data clean rooms, brands can bring in impression level data from walled gardens (like Netflix's new ad platform or YouTube) into a secure BigQuery environment to match with their first party data without violating privacy. Google's own Ads Data Hub is an example that works with BigQuery to allow advertisers to attribute conversions to YouTube/Google ads by querying data within a privacy safe environment. Netflix, as noted, is enabling similar capabilities via Snowflake and clean room partners[43]. All these tools point to one trend: using large scale data processing to reconstruct the customer journey, even when pieces are missing or when attribution must be probabilistic rather than deterministic.

In practice: A company implementing a long term attribution framework will often use GA4 and other tools for real time tracking, but ultimately funnel all that data into BigQuery for storage and advanced analysis[12]. They would design a schema to capture each touchpoint with fields like user_id, timestamp, channel, campaign, etc. Over time, as data accumulates, they could ask complex questions such as: *“Show me all users who made a purchase in 2025 that had first interacted with our brand back in 2023 or earlier.”* Or *“What fraction of users who see our YouTube ads eventually convert, even if the conversion happens 6 months later?”* Those questions can be answered by joining YouTube impression logs (via Ads Data Hub) with internal sales data in BigQuery. Another real example: Google's marketing team often cites using BigQuery with GA4 to do process mining of customer journeys – essentially mapping out all paths users take and finding common sequences leading to conversion[44][45]. This helps find “hidden” journeys that standard funnels miss. In sum, BigQuery is the sandbox where all data converges. It's where you can perform retroactive analyses, test different attribution models, and even simulate long term outcomes (like predicted lifetime value) by looking at years of user behavior. It requires data engineering and analytics expertise, but it unlocks attribution analysis that is simply not possible within individual marketing platforms alone.

Advanced Techniques: Probabilistic Modeling and Server Side Tracking

To truly capture fragmented or long latency purchase journeys, brands are employing two key techniques across all platforms: probabilistic models and server side tracking.

Probabilistic Attribution Models: When deterministic tracking (tying a conversion to a known click or user ID) fails or is incomplete, probabilistic models step in to estimate the connection. We've already touched on several examples: GA4's conversion modeling for users who decline cookies[6], Meta's modeled conversions for iOS users[28], and MMM in BigQuery. Another probabilistic approach is using Markov chain modeling on user paths (common in multi touch attribution studies). This calculates the likelihood of conversion with and without each touchpoint, thereby assigning a probabilistic "credit" to every touch based on its contribution to conversion probability. One might find, for example, that if Touch A (say, a product video view) is removed, the probability of eventual purchase drops by 10% – thus Touch A gets 10% attribution credit. Such a model can be run on journey data exported to BigQuery or using specialised tools, and it helps reconstruct the influence of earlier touches that might not directly get credit in a last click model. Salesforce's guidance explicitly mentions considering algorithmic attribution and even custom models tailored to your business[46][47]. The future of attribution is likely a blend of approaches: deterministic where possible, and probabilistic elsewhere. Brands like Netflix rely heavily on incrementality testing – a probabilistic method where you expose a test group to a marketing touch and compare behavior to a control group. Netflix's content acquisition study essentially did this to infer how much a new show contributed to subscriber sign ups[42]. Even if you can't track a specific user from seeing a billboard to subscribing, these models give a *statistical linkage* between the two events. Bottom line: probabilistic models are crucial when direct tagging is impossible (e.g. offline ads) or blocked (privacy settings). They fill in the puzzle pieces so that early marketing efforts still get some recognised credit in a long customer journey[48].

Server Side Tracking (and Tagging): One practical challenge in journey tracking is data loss due to cookies being blocked, ads being skipped, or app tracking opt outs. Server side tracking is a technique where events are sent from your own server (or cloud tag manager) rather than the user's device, which can improve data quality and persistence. For example, using Google Tag Manager Server Side or GA4's Measurement Protocol, a brand can transmit page views or conversions directly to GA4 from a server with first party context (your own domain). This can bypass cookie expiration or browser restrictions, ensuring a user's persistent ID (e.g. a login ID or a server set cookie) stays intact longer. It also allows you to attach additional data (like a CRM customer ID or lifetime value segment) to each event in a secure way[49]. Many businesses now send purchase events *both* from the client and from the server – the server copy might include a robust user identifier and be less likely to be blocked. Similarly, we discussed Meta's Conversions API, which is essentially server side event tracking for Facebook. By 2025, server side tagging has become a best practice for long term attribution because it grants more control: you decide how to persist identifiers (for example, setting a first party cookie that lasts 1+ years) and you can forward events to multiple endpoints (analytics, ads, CRM) reliably. This technique was highlighted in a 2023 industry paper which stated: only integrated ecosystems of first party data *and server side tracking* can meaningfully connect early impressions to future conversions[3][49]. The caveat is that moving tracking server side doesn't eliminate privacy obligations – you must still obtain user consent for tracking and data sharing. But it can reduce dependency on fragile client side storage and link users across sessions more effectively. As an example, suppose a user clicks a Google Ad but doesn't allow third party cookies. With server

side tagging, your server can still read a Google Click ID from the landing page URL and store it in a first party cookie. If the user converts much later, your server can send that Click ID with the conversion, allowing Google Ads (or GA4) to attribute the conversion to the original ad despite the long delay. Without server side capture, that link might have been lost after the browser's cookie reset. Thus, server side tracking acts as a safety net for data continuity in the customer journey.

In summary, combining probabilistic models and server side tracking yields a powerful approach to reconstructing customer journeys that are otherwise broken or obscured. Probabilistic modeling gives you insight and attribution credit when user level linking isn't possible, and server side tracking maximises the user level data you *can* collect by working around technical limitations. Brands implementing long term attribution should deploy both: use server side methods to log as much high quality, consented data as possible, and layer on statistical models to infer the rest.

Key Takeaways and Recommendations for Long Term Attribution

Designing a robust long term attribution framework requires strategy across people, process, and technology. Here are practical recommendations for brands looking to tag customer journeys from the first impression to the eventual conversion, no matter how long it takes:

Establish Persistent User IDs: Encourage or incentivise users to log in, register, or join loyalty programs early in their journey. A persistent identifier (email, account ID, loyalty number) is the backbone of long term tracking[49]. For example, Starbucks' success is tied to its Starbucks Rewards ID being used across mobile, web, and in store – linking all those interactions. If logins aren't feasible, use first party cookies with long expiration and unify cross device data via techniques like hashed emails or device graphs. *The goal is to have a stable key to join data over time.*

Instrument Every Touchpoint (with Context): Deploy comprehensive event tagging on all marketing and product channels. Use tools like GA4, Pixels, and CRM campaign records to log each impression, click, site visit, app action, and offline touch. Tag URLs with UTM parameters or other campaign codes from the very start of a campaign[22] – this ensures that when a user does convert later, you have stored the origin attribution in their history. Capture timestamps and identifiers with each event so you can sequence and connect them later. Essentially, leave a breadcrumb trail for every customer interaction.

Integrate and Store Data in a Unified Repository: Break down data silos by feeding all touchpoint data into a central warehouse or customer data platform. Connect your web/app analytics, ad platforms, email systems, and CRM. This may involve using APIs, nightly data pipelines, or cloud connectors. By consolidating data in BigQuery or a similar store, you enable cross channel identity resolution and unlimited lookback analysis[12]. Plan for long data retention – keeping multiple years of granular data will let you observe truly long term behaviors. Investing in data integration and governance up front pays off when you can actually *find* and trust the data linking an impression from 3 years ago to a purchase today[39].

Apply Multi Touch Attribution Models (and Evolve Them): Adopt a multi touch mindset – don't default to last click or assume a single touchpoint "caused" the conversion. Use the attribution features in GA4, Meta, or Salesforce Campaign Influence to get immediate multi touch insights

(e.g. GA4's data driven model, Salesforce's first/last touch reports). For deeper analysis, leverage custom models: test out linear, time decay, position based, or algorithmic models and see which aligns best with your business reality[50][46]. The "right" model may change over time; regularly compare attribution outputs with actual business outcomes (like customer LTV) and refine accordingly[51][52]. *Remember:* attribution modeling is not one and done – treat it as an iterative process. Also, consider incrementality testing (hold out tests) for validation: e.g. pause a channel in one region and see if conversions drop, to verify that your modeled attribution wasn't over crediting that channel.

Leverage Predictive Metrics for Long Term Value: Traditional ROAS (Return on Ad Spend) or CPA metrics might undervalue early funnel touches that pay off later. Augment your framework with long term value metrics. For instance, track customers acquired from a certain campaign and measure their 12 month value or predicted lifetime value (pLTV)[53]. Optimise campaigns not just for immediate conversion, but for those that bring in high value or repeat customers over time. This shift in KPIs (as Gymshark did by focusing on user conversion and retention) encourages marketing efforts that build brand equity and loyalty, which often manifest in delayed conversions. In your data models, you can allocate credit not just for the immediate sale, but for the *future value* a touchpoint generates (e.g. an impression that led to a subscription that renews monthly).

Implement Server Side Tracking with Consent: Improve your data capture by sending events server side (using tools like GTM Server Side, Conversions API, or custom webhooks) to extend attribution windows and reliability[49]. Server side methods allow you to maintain consistent user identification (via first party cookies or login tokens) beyond the limits of browser cookies, and to enrich events with CRM data (for example, tagging a conversion with a user's segment or tenure). However, ensure you have user consent for the data you collect; be transparent in your privacy policy about long term tracking. Use features like GA4's Consent Mode if applicable, which works with modeling to respect user choices while still filling measurement gaps.

Monitor and Adapt to Privacy Changes: The landscape of tracking is continually evolving (browser restrictions, regulations like GDPR/CCPA, and platform policies). Build your long term attribution framework to be resilient by focusing on first party data (which you directly collect with permission) and avoiding overly fingerprinting based methods that might be regulated against. Embrace privacy centric solutions such as aggregated measurement (e.g. Google's Aggregated Conversions or Meta's Aggregated Event Measurement) which provide conversion data in compliant ways. These solutions often involve modeling – be prepared to explain internally that some reported conversions will be *modeled/estimated*, not exact. Educate stakeholders that modeled data is a feature, not a flaw, in the new era of attribution[28][54].

Case Study and Learn: Look at brands who are successfully doing this. Starbucks, for example, connects data from mobile app orders, loyalty scans in store, and digital campaigns to create personalised marketing – they can target a user with an offer on the app and see the sale in store weeks later, all tied to that user's ID[23]. Gymshark's analytics revamp shows the value of focusing on users over sessions, leading to concrete UX improvements and higher conversion rates[17][18]. Netflix's data science approach underscores the importance of measuring the *incremental impact* of content and marketing over time, not just immediate clicks[42]. These examples reinforce that long term attribution is achievable with the right mix of data and tools –

and that it yields actionable insights (like which early touchpoints truly drive valuable customers).

In conclusion, tracking seemingly “random” or delayed purchases back to their original marketing touchpoints is a challenging but solvable puzzle. By deploying an integrated tech stack (GA4, Meta, Salesforce, BigQuery, etc.), persisting customer identifiers, and layering in advanced modeling, brands can connect the dots from the first impression to the final conversion – even if those dots are spread across years and channels. The process hinges on building a unified, privacy conscious data foundation and constantly refining attribution assumptions with real data. Those who succeed will gain a deep understanding of the *entire* customer journey and be able to invest in marketing strategies that pay off in the long run. As one report summarised: “*Tracking an impression that leads to a sale five years later is challenging but feasible. A unified ecosystem linking CRM, analytics, and server side data allows brands to attribute success to early touchpoints*”[55]. By following the practices above, marketing teams can move beyond short term metrics and start optimising for enduring customer relationships and lifetime value, confidently proving the ROI of every touchpoint along the way.

Sources:

Google Analytics 4 cross device and identity documentation[56][5]

InfoTrust: Cross Device Tracking in GA4 – identity spaces and blended stitching[4][57]

Think with Google – Gymshark Case Study: first party data focus and user level analytics outcomes[7][8]

Jon Loomer, *Meta Ads Attribution Guide*: default windows and cross device matching via Meta accounts & API[25][20]

Wpromote, *Facebook’s Modeled Conversions*: ML filling in conversions for iOS and cross device gaps[28][29]

Salesforce, *Multi Touch Attribution Best Practices*: cross device tracking and matching techniques[33], integrating offline touchpoints and using ML for attribution[37]

OptiMine blog, *Cross Channel Marketing (Starbucks example)*: loyalty program as cross channel ID and its impact[23][24]

Netflix Research – *Incremental Acquisition Impact of Content Launches*: advanced modeling of content (marketing) influence on sign ups[42]

User Provided Research (Bushell, 2025): *Tracking Impressions Across the Full Funnel – Long Term Attribution Framework*[49][22][55] (synthesising industry best practices for persistent IDs, server side tracking, data warehousing, and probabilistic attribution models).

[1] [3] [12] [22] [32] [48] [49] [53] [55] Long_Term_Attribution_Framework_ChrisBushell.docx
file:///file 6sWwsJV62uhiMsNBDbn3NG

[2] [33] [37] [38] [39] [40] [46] [47] [50] [51] [52] Multi Touch Attribution: What It Is & Best Practices | Salesforce US

<https://www.salesforce.com/marketing/multi touch attribution/>

[4] [5] [6] [9] [10] [13] [56] [57] Cross Device Tracking in Google Analytics 4: How to Identify Users Across Multiples Devices

[https://infotrust.com/articles/cross device tracking in google analytics 4/](https://infotrust.com/articles/cross-device-tracking-in-google-analytics-4/)

[7] [8] [14] [15] [16] [17] [18] [41] Lessons from Gymshark's data transformation Think with Google

[https://www.thinkwithgoogle.com/intl/en-emea/future of marketing/digital transformation/gymshark data transformation/](https://www.thinkwithgoogle.com/intl/en-emea/future-of-marketing/digital-transformation/gymshark-data-transformation/)

[11] [19] [20] [21] [25] [26] A Complete Guide to Meta Ads Attribution Jon Loomer Digital

[https://www.jonloomer.com/meta ads attribution/](https://www.jonloomer.com/meta-ads-attribution/)

[23] [24] What is Cross Channel Marketing? OptiMine

[https://optimine.com/blog/what is cross channel marketing/](https://optimine.com/blog/what-is-cross-channel-marketing/)

[27] About Cross Device Reporting | Meta Business Help Center

<https://www.facebook.com/business/help/266118906912295>

[28] [29] [30] [54] How to Unlock Better Measurement with Facebook's Modeled Conversions

[https://www.wpromote.com/blog/analytics/facebook's modeled conversions](https://www.wpromote.com/blog/analytics/facebook-modeled-conversions)

[31] About Meta's modeled conversions | Meta Business Help Center

<https://www.facebook.com/business/help/311705270326952>

[34] Multi touch attribution in Salesforce Do I need Pardot? Reddit

https://www.reddit.com/r/salesforce/comments/uc9hxx/multitouch_attribution_in_salesforce_do_i_need/

[35] Multi Touch Attribution Dashboard Salesforce Help

https://help.salesforce.com/s/articleView?id=mktg.pardot_b2bma_multitouch_attribution_dashboard.htm&language=en_US&type=5

[36] How Customisable Campaign Influence Works Salesforce Help

https://help.salesforce.com/s/articleView?id=sales.campaigns_influence_customisable_understanding.htm&language=en_US&type=5

[42] Estimating Incremental Acquisition of Content ... Netflix Research

[https://research.netflix.com/publication/estimating incremental acquisition of content launches in a subscription](https://research.netflix.com/publication/estimating-incremental-acquisition-of-content-launches-in-a-subscription)

[43] Netflix Gets Into Clean Rooms; Can AI Fix CTV's Fill Rate Problem?

[https://www.adexchanger.com/daily news roundup/monday 26082024/](https://www.adexchanger.com/daily-news-roundup/monday-26082024/)

[44] Process Mining Your Customer Journey Using GA4 & BigQuery

[https://measuremindsgroup.com/customer journey using ga4 bigquery](https://measuremindsgroup.com/customer-journey-using-ga4-bigquery)

[45] Brewing Success: A Starbucks Case Study in Data Driven Customer ...

[https://medium.com/@senatorlionel2/brewing success a starbucks case study in data driven customer engagement and how your business ac1bbc78cf67](https://medium.com/@senatorlionel2/brewing-success-a-starbucks-case-study-in-data-driven-customer-engagement-and-how-your-business-ac1bbc78cf67)