

Défier les limites de GA-4



Comment franchir les limites de Google Analytics 4?

Dans le monde actuel axé sur les données, les entreprises doivent clairement comprendre les besoins de leurs clients, afin d'avoir un avantage concurrentiel. L'analyse web constitue un élément important favorisant cette compréhension, car elle fournit des informations précieuses sur la façon dont les utilisateurs interagissent avec les actifs numériques d'une entreprise.

Google Analytics 4 (GA4) est la dernière version de la plateforme d'analyse web de Google. Il présente plusieurs nouvelles fonctionnalités et changements par rapport à son prédécesseur, Universal Analytics (UA). GA4 dispose d'un [modèle de données basé sur les événements](#), des capacités de suivi améliorées et une intégration plus avancée dans l'écosystème de Google.

Le passage à un modèle de données moderne basé sur les événements est le changement le plus significatif et nécessite une rééducation et une planification ; cependant, un autre changement qu'il est important de ne pas négliger pourrait affecter substantiellement votre entreprise : Comment GA4 prend-il en charge le suivi côté serveur ?

Avantages du suivi côté serveur

Avec l'évolution du paysage analytique, il est de plus en plus difficile de collecter des données complètes en ce qui concerne les conversions et l'attribution. Cette évolution s'explique en grande partie par deux facteurs :

- vitesse de la page - Il est estimé que chaque seconde de retard dans le chargement d'une page sur ordinateur ou téléphone mobile réduit les taux de conversion de 20 à 22 % ;
- bloqueurs de publicité - Plus de 40 % des personnes dans le monde utilisent désormais des bloqueurs de publicité.

Si vous ne comptez que sur le suivi côté client, il est probable que votre taux de conversion en pâtisse et que vous perdiez une quantité importante de données précieuses au profit des bloqueurs de publicité. Le suivi côté serveur peut vous aider à rester compétitif face à ces défis, car il favorise:

- l'accélération des performances du site web : le suivi côté serveur réduit le nombre de code nécessaires pour charger le côté client, ce qui peut entraîner des améliorations considérables des mesures de vitesse des pages ;
- la résilience face aux bloqueurs de publicité : le suivi côté serveur vous permet de collecter des données malgré les bloqueurs de publicité, ce qui vous permet de récupérer d'importants volumes de données autrefois indisponibles.

Plus important encore, le suivi côté serveur vous permet de vous approprier votre collecte de données de première partie. Plutôt que d'insérer du code tiers dans l'ensemble du site, vous pouvez envoyer des données côté serveur afin d'exercer un contrôle plus granulaire sur ce qui se passe avec les données.

Défis liés à GA4

Par rapport à Universal Analytics, le suivi côté serveur de GA4 est moins indépendant et nécessite le soutien du suivi côté client pour activer des fonctions essentielles dont vous pouvez dépendre. Les fonctionnalités qui nécessitent un suivi côté client sont : l'attribution, la sessionnalisation, la géolocalisation et les conversions pour le remarketing Google Ads.

Il s'agit là d'un compromis douloureux. Si vous ne comptez que sur le suivi côté serveur, vous n'aurez pas accès à des fonctionnalités principales. Si vous ne comptez que sur le suivi côté client, les performances de votre site en pâtiront et vous ferez face à des pertes de données excessives, car le suivi côté client est notoirement sensible aux bloqueurs de publicité.

Le présent guide propose deux étapes pour relever ces défis. En adoptant une approche hybride face à GA4 à l'aide de RudderStack pour activer facilement le suivi côté serveur et côté client, vous pouvez obtenir le meilleur des deux options et débloquer toute la puissance de GA4. À terme, vous souhaitez peut-être aller au-delà de GA4 pour répondre à des questions plus difficiles. Pour les équipes intéressées par cette deuxième étape, nous aborderons comment mettre en œuvre une stratégie d'analyse d'entrepôt pour aider votre entreprise à mieux s'approprier les données, à pérenniser vos analyses et à améliorer l'intégration avec le reste de votre pile technologique.

Options du suivi côté serveur de GA4

GA4 propose deux options pour le suivi côté serveur : Google Tag Manager (GTM) côté serveur ou l'API du protocole de mesure GA4. Nous détaillons ci-dessous les deux options et leurs insuffisances.

Google Tag Manager côté serveur

Le [balisage côté serveur de GTM](#) de GA4 est différent de celui d'un conteneur GTM du web. Il est hébergé sur un conteneur côté serveur (GCP ou autre), ce qui permet de déplacer l'instrumentation hors du navigateur, mais il s'accompagne de quelques insuffisances importantes :

- il ne s'agit pas d'un « pur côté serveur » sans code de navigateur, car il faut toujours un conteneur GTM du web pour envoyer des événements au conteneur côté serveur ;
- les balises du navigateur de GA4 sont utilisées comme une couche de données côté serveur, et les données sont transformées en d'autres balises (Facebook, Snapchat, etc.). Ce mappage des données peut présenter des points de défaillance supplémentaires pour certains sites. Si votre équipe de développement web par exemple, entretient la couche de données de GTM pour le suivi des événements, GTM côté serveur ne peut pas récupérer ces données en mode natif. Au lieu de cela, GTM côté serveur récupère les données qui passent par le balisage de GA4, ce qui peut présenter des divergences par rapport à l'extraction à partir de la couche de données brutes.

API du protocole de mesure de GA4

L'API du protocole de mesure vous permet d'écrire des codes et d'envoyer des données directement à <https://www.google-analytics.com/mp/collect>. C'est un moyen efficace d'augmenter les données collectées côté client, et c'est la méthode utilisée par RudderStack pour envoyer les données côté serveur GA4, mais sa mise en œuvre standard est problématique :

- Les rapports d'attribution prêts à l'emploi de GA4 ne prennent pas en charge les données côté serveur, ils nécessitent un suivi du navigateur ;
- Le remarketing de Google Ads nécessite un suivi du navigateur ;
- Les informations géographiques prêtes à l'utilisation nécessitent un suivi du navigateur ;
- La sessionnalisation nécessite un identifiant de session, un identifiant de client et un horodatage qui correspondent à une session particulière, et cette mise en œuvre nécessite une assurance qualité minutieuse ;
- La mise au point d'un identifiant client côté serveur pour les utilisateurs anonymes peut nécessiter un nouveau travail de développement pour certaines entreprises.

Actuellement, le fait de n'utiliser que l'API du protocole de mesure pour la collecte des données élimine des fonctionnalités essentielles des rapports de GA4. Le balisage côté serveur de GTM permet de contourner ce problème, car il nécessite toujours un conteneur web ou une balise Google (gtag.js) codée dans le site, mais, comme indiqué, cette approche comporte ses propres limites.

Les deux approches côté serveur présentant des inconvénients significatifs, la plupart des entreprises auront besoin d'une mise en œuvre en mode hybride de GA4 qui comprend le suivi côté serveur et la prise en charge par le code du navigateur.

Architecture d'un déploiement hybride

Le déploiement hybride permet d'obtenir le meilleur des deux options tout en glanant un ensemble plus complet d'informations. Principaux avantages du déploiement hybride :

- L'attribution peut être créée et mise à jour via le code du navigateur (GTM, gtag.js) ;
- La sessionnalisation peut être générée par le code du navigateur (GTM, gtag.js) ;
- Les données de géolocalisation peuvent être glanées par le code du navigateur (GTM, gtag.js) ;
- Les conversions pour le Remarketing Google Ads peuvent être enregistrées par le code du navigateur (GTM, gtag.js).

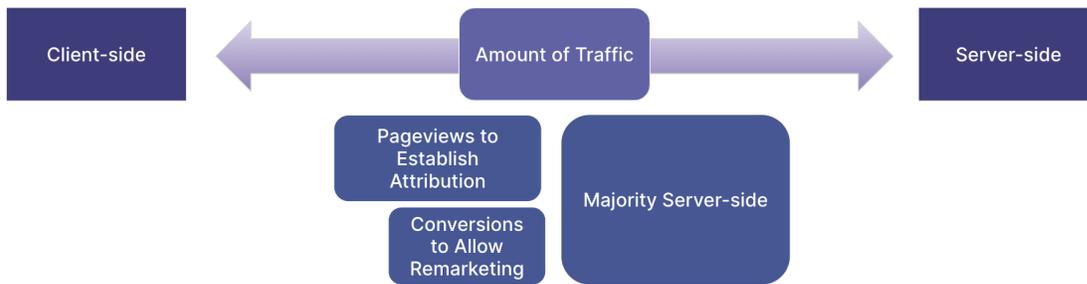
Pour de nombreuses entreprises, ces avantages ne sont pas négociables parce qu'ils permettent à leurs équipes de disposer de capacités sur lesquelles elles comptent. Les déploiements hybrides présentent toutefois certains inconvénients qu'il convient de prendre en compte. Les déploiements hybrides sont plus chronophages en ce qui concerne leur mise en place, ils nécessitent toujours du code côté client, ce qui affecte les performances, et ils sont partiellement sensibles aux bloqueurs de publicité.

Il existe trois façons principales de mettre en place un déploiement hybride, et chaque option implique ses propres compromis. Nous allons l'illustrer par quelques exemples ci-dessous.

Option 1 : Côté serveur avec côté client si nécessaire

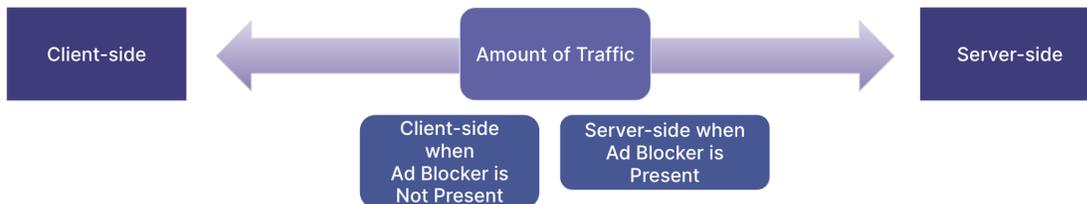
Ce déploiement utilise le suivi côté serveur partout où le suivi côté client n'est pas nécessaire pour l'attribution, le remarketing ou la géolocalisation des données. Il a le moins d'impact sur les performances côté client et réduit la perte de données due aux bloqueurs de publicité, mais il

exige un engagement organisationnel important en faveur du suivi côté serveur, qui nécessite une formation et une mise à jour.



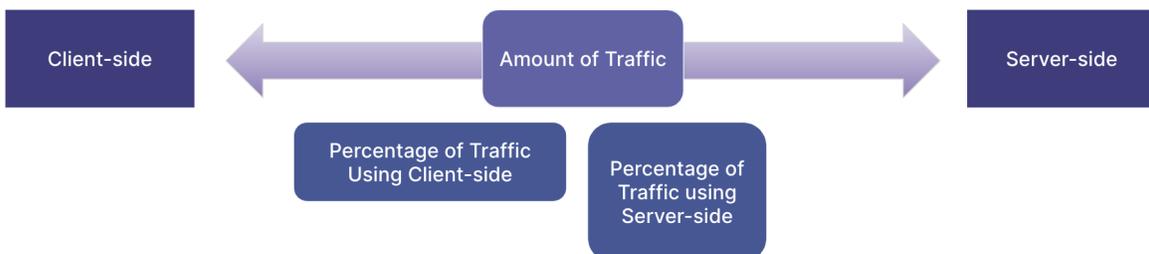
Option 2 : Repli côté serveur pour les bloqueurs de publicités

Dans cette configuration, le suivi côté serveur n'est utilisé que lorsqu'un bloqueur de publicité côté client est détecté. Elle réduit la perte de données due aux bloqueurs de publicité, mais a toujours des impacts côté client sur les performances.



Option 3 : Séparer le côté serveur et le côté client

La séparation du suivi côté serveur et côté client permet d'exécuter les deux méthodes de suivi en parallèle. Les données sont ensuite dédoublées dans les rapports. Lorsqu'une réponse positive côté client ne peut être trouvée pour une réponse positive côté serveur spécifique, l'impact positif des performances côté serveur est démontré.



Cette approche permet aux entreprises de comparer directement la collecte de données côté client à la collecte de données côté serveur, et comme le suivi côté client est toujours pleinement opérationnel, vous disposez de plus de temps pour former votre équipe et mettre à jour le suivi côté serveur. Cependant, comme vous pouvez le constater, cette approche nécessite des ressources de rapport pour la déduplication et a toujours un impact sur les performances du côté du client.

Avec GA4, une mise en œuvre hybride est le seul moyen d'accéder à un ensemble de données d'attribution plus complet, plus unifié et plus précis, avec le moins d'impact possible sur les performances. Cependant, la mise sur pied d'une implémentation hybride avec les outils fournis par Google est compliquée et fait appel à de la maintenance. C'est pourquoi nous avons ajouté une option de déploiement hybride à l'intégration de GA4 de RudderStack.

Utilisation du mode hybride RudderStack

RudderStack est une plateforme de données clients native de l'entrepôt qui facilite le transfert direct des événements comportementaux de l'utilisateur vers et à partir de votre entrepôt de données. En outre, elle fournit une bibliothèque d'intégration qui permet de transmettre des données à plus de 180 autres outils. Ces outils comprennent des plateformes d'analyse populaires telles que Google Analytics 4.

La destination GA4 de RudderStack prend en charge à la fois une connexion en mode cloud et une connexion en mode appareil. Nous allons ici mettre en évidence les principales différences entre les deux.

Le mode appareil charge le SDK GA4 (gtag) directement sur un site web. Il est conçu pour les connexions côté client où les événements ne passent pas par le serveur de RudderStack. Ce type de connexion prend en charge l'ensemble des fonctionnalités disponibles dans GA4, mais il est vulnérable face aux bloqueurs de publicité et peut avoir un impact important sur les performances du site.

 data:image/svg+xml;...	200	svg+xml
 rudder-analytics.min.js	200	script
 ?p=cdn&v=2.32.0&writeKey=2HxrdF4Y9oKmM2pmjQmTB9i...	200	xhr
 page	200	xhr
 GA4.min.js	200	script
 js?id=G-TVZYF9K3L4	200	script
 js?id=G-Q56X1VG5L6&l=dataLayer&cx=c	200	script

Le mode Cloud envoie des requêtes à l' [API du protocole de mesure](#), qui représente la connexion côté serveur de GA4. Cela permet aux développeurs d'envoyer directement des pages vues et d'autres événements d'interaction utilisateur aux serveurs GA4 via des requêtes HTTP, permettant ainsi le suivi du comportement de l'utilisateur sur diverses plateformes, y compris les sites Web et les applications mobiles, ainsi que les événements hors ligne qui ne sont pas générés au sein d'une application Web. Contrairement au mode appareil, les événements du mode cloud sont traités par le serveur RudderStack et peuvent être visualisés dans la visionneuse d'événements en direct de RudderStack.

The screenshot displays the RudderStack event viewer interface. At the top, it shows 'ga4 - cloud mode live events' with a timestamp of '04:39' and buttons for 'Pause' and 'Clear Events'. Below this, there are tabs for 'All', 'Success', and 'Failed'. A search bar and filter options are visible. The main area is divided into two columns: 'EVENT' and 'ERROR MESSAGE'. The 'EVENT' column lists several events, with the last one, 'button_clicked', highlighted in a light red box. The 'ERROR MESSAGE' column shows a right-pointing arrow for each event. To the right, a 'PAYLOAD' section shows the JSON data for the selected event, including details like 'name', 'params', 'client_id', and 'timestamp_micros'.

EVENT	ERROR MESSAGE
page 4/27/2023, 5:51:58 PM	→
page 4/27/2023, 5:51:58 PM	→
button_clicked track 4/27/2023, 5:52:00 PM	→
button_clicked track 4/27/2023, 5:53:39 PM	→
button_clicked track 4/27/2023, 5:53:39 PM	→

```
8     "name": "button_clicked",
9     "params": {
10      "page": "Home",
11      "context": "Blog",
12      "service": "Research",
13      "session_id": "1682631961472",
14      "engagement_time_msec": 1
15    }
16  },
17  ],
18  "client_id": "405b6fe8-5eb7-4340-8fc2-1ba7e29b0878",
19  "timestamp_micros": 1682632414000000
20  },
21  "JSON_ARRAY": {}
22  },
23  "type": "REST",
24  "files": {},
25  "method": "POST",
26  "params": {
27    "api_secret": "TLzILNGAQKeXBYJ2M9GKXw",
28    "measurement_id": "G-QCDZBIKQYV"
29  },
30  "userId": "",
31  "headers": {
32    "HOST": "www.google-analytics.com",
33    "Content-Type": "application/json"
34  },
35  "version": "1",
36  "endpoint": "https://www.google-analytics.com/mp/collect"
37 }
```

La connexion en mode cloud offre les avantages du suivi côté serveur décrits ci-dessus, mais comme indiqué, les principales fonctionnalités de GA4 ne prennent pas en charge les données côté serveur, donc, s'appuyer uniquement sur le mode cloud pour les rapports d'analyse ne permet pas d'obtenir une image complète et entrave les fonctionnalités de GA4. Google déclare même que le suivi côté serveur est destiné à compléter la configuration gtag côté client, et non à la remplacer complètement.

Pour permettre aux équipes de bénéficier des avantages des deux modes de connexion, l'intégration de GA4 par RudderStack inclut désormais un troisième mode de connexion : le mode hybride. Cette nouvelle fonctionnalité permet un déploiement hybride de GA4 en une seule étape qui combine automatiquement les informations provenant des connexions côté client et côté serveur.

Ce qui se passe en mode hybride :

Le choix du Mode hybride divise les événements en deux flux :

- le mode appareil enregistre les événements de page ;
- le mode Cloud enregistre les événements de suivi et d'identification.

Dans le backend de RudderStack, nous récupérons l'identifiant de la session et l'identifiant du client à partir des événements gtag envoyés par le mode appareil et nous les fusionnons avec les événements enregistrés par le mode cloud afin d'établir un rapport précis et complet des interactions de l'utilisateur au cours de la session.

Cette configuration améliore la vitesse du site, car elle réduit les appels réseau en enregistrant uniquement les appels de page avec le mode appareil (côté client) qui prend en charge le remarketing, l'UTM et le suivi de la géolocalisation, tandis que les appels de suivi contenant des données cruciales sont enregistrés plus rapidement et avec moins de perte de données en tirant parti du mode Cloud (côté serveur).

Avantages du mode hybride

La connexion [Mode hybride de GA4](#) de RudderStack vous permet d'accéder à une implémentation qui offre le meilleur des deux options. Vous obtiendrez un ensemble de données d'attribution plus complet, plus unifié et plus précis avec le moins d'impact possible sur les performances du front-end grâce à une combinaison de suivi côté client et côté serveur. Cette fonction suit l'approche décrite ci-dessus dans la première option - l'option côté serveur avec côté client selon les besoins.

Voici un exemple d'une configuration en mode hybride de GA4 fonctionnant sur un site web :

The screenshot shows a browser's developer console with the following details:

- Network Panel:** A list of requests is shown. The request for `GA4.min.js` is highlighted with a red box and labeled '1'. The request for `collect?v=2&tid=G-Q56X1VG5L6>m=45je3510&p=17107...` is highlighted with a red box and labeled '2'.
- Request Details Panel:** The 'Request URL' field is highlighted with a red box and labeled '3'. The URL is `https://www.google-analytics.com/g/collect?v=2&tid=G-Q56X1VG5L6>m=45je3510&p=1710798314&cid=21c4a4a4-e41f-48e8-933a-24d133dffab1&ul=en-us&sr=1&12x982&uaa=arm&uab=64&uafvl=Chromium%3B112.0.5615.137%7CGoogle%2520Chrome%3B112.0.5615.137%7CNot%253AA-Brand%3B99.0.0.0&uamb=0&uam=&uap=ma05&uapv=13.3.1&uaw=0&_eu=AEA&_s=1&sid=1683053908405&sct=7&seg=1&dl=https%3A%2F%2Fwww.saramashfej.com%2F&dt=Sara%27s%20Demo%20Site&en=scroll&ep.debug_mode=true&epn.percent_scrolled=90&_et=3`. The 'Request Method' is POST, the 'Status Code' is 204, the 'Remote Address' is 142.251.41.46:443, and the 'Referrer Policy' is strict-origin-when-cross-origin.

1. Fait référence au gtag chargé sur le site : il s'agit de la connexion côté client.
2. Demande adressée à l'API du protocole de mesure : il s'agit de l'appel effectué côté serveur.
3. L'URL de la requête qui représente le point de terminaison du protocole de mesure vers lequel les événements seront envoyés : <https://www.google-analytics.com/g/collect> along with the query parameters.

Cette fonctionnalité facilite également la mise en œuvre d'un déploiement en mode hybride de GA4. Grâce à l'intégration de RudderStack, il n'est pas nécessaire de créer deux connexions distinctes, puis de dédoubler les événements par filtrage et transformation. Il suffit d'établir une connexion et d'activer le mode hybride.

[Consultez la documentation](#) et cet article fondamental sur [la façon d'envoyer des données au mode hybride de GA4](#) pour apprendre à utiliser le mode hybride de GA4 de RudderStack.

Avec Rudderstack, vous pouvez tirer le meilleur parti de GA4, mais pour des analyses plus approfondies, vous risquez de trouver GA4 limité. Son modèle de base ne permet pas une grande flexibilité ni une grande appropriation des données, et il y a tout simplement des questions auxquelles vous ne pouvez pas répondre avec uniquement GA4.

Heureusement, avec RudderStack, lorsque vous vous heurtez aux limites fondamentales de GA4, vous disposez déjà de la couche de données nécessaire à l'analyse de l'entrepôt. Grâce à l'analyse d'entrepôt, vous pouvez faire des analyses pertinentes qui fournissent de meilleures informations sur l'ensemble du parcours de l'utilisateur et qui sont cohérentes dans l'ensemble de la pile de données.

Aller au-delà de GA4

De nombreuses équipes chargées des données et du marketing vont au-delà de Google Analytics. GA4 est meilleur que UA à bien des égards, mais en raison de son modèle de base, des insuffisances persistent même avec une mise en œuvre hybride.

Les avantages d'être propriétaire de son infrastructure d'analyse

Heureusement, vous n'avez pas à soumettre votre équipe chargée de données, ou toute autre équipe, aux limites de Google Analytics. En utilisant RudderStack pour la collecte de données de première partie et votre entrepôt pour le stockage, la modélisation et le service des données,

vous pouvez élaborer une pile analytique à l'épreuve du temps qui fournira des données auxquelles vous pouvez faire confiance et qui évoluera facilement avec votre entreprise.

« Les données ont changé, et pour ceux qui dépendent encore de Google Analytics, c'est le moment idéal pour moderniser votre pile de données. Les données de vos clients constituent l'actif le plus précieux de votre entreprise, et vous devez posséder toutes les données et cesser de dépendre des outils tels que GA pour les héberger ou les rapporter. Avec un CDP solide, vous pouvez désormais suivre toutes vos données clients et les acheminer vers un entrepôt, tout en étant propriétaire de vos données. À l'avenir, les entreprises qui s'approprient leurs données seront celles qui gagneront. »

Dan McGaw

Fondateur et PDG de McGaw.io, experts en technologie et en analyse

L'analyse avec une infinité d'options

En séparant votre infrastructure de prise des données de vos couches de stockage et de transformation des données, ainsi que des outils d'analyse eux-mêmes, vous avez la possibilité de modifier n'importe quel composant individuel de votre pile à mesure que les besoins d'analyse de votre entreprise changent.

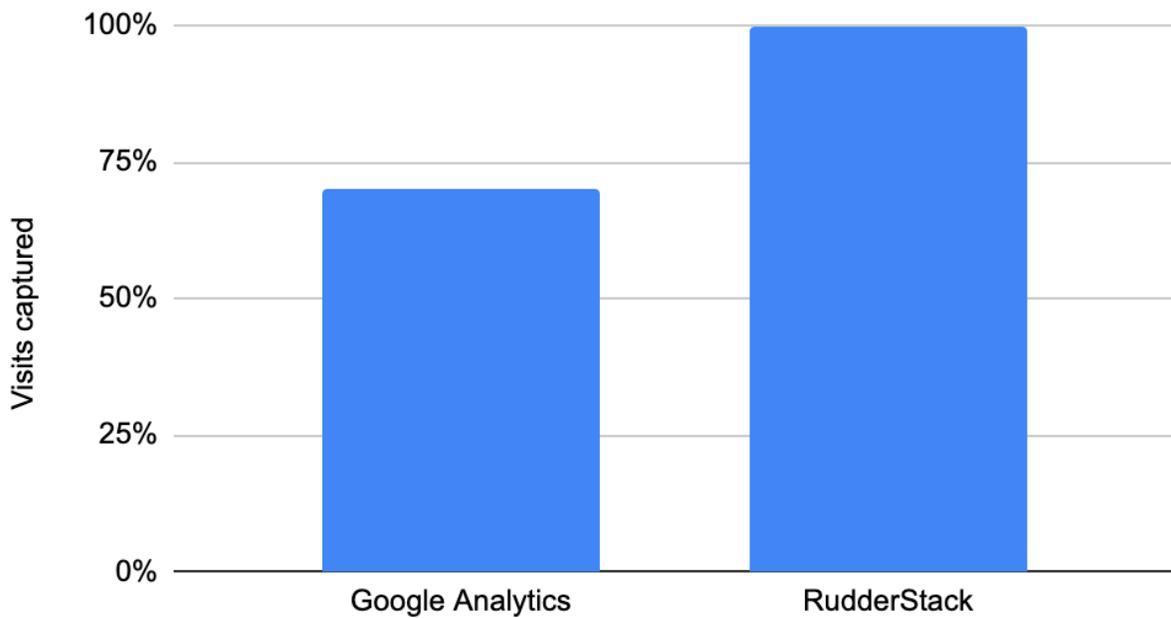
Vos équipes chargées des tests A/B ou des produits souhaitent essayer un nouvel outil d'analyse ? Pas de problème : il suffit de diriger le flux d'événements vers une nouvelle destination sans ouvrir de tickets de développement (nous y reviendrons). Vos analystes doivent-ils modifier les mesures pour tenir compte d'un nouveau plan de tarification ? Vous pouvez évoluer rapidement en mettant à jour les requêtes exécutées sur votre entrepôt sans toutefois revoir complètement l'instrumentation des événements et l'ensemble de votre couche de visualisation. Que se passe-t-il si votre équipe scientifique en charge des données élabore un nouveau modèle de recommandations sur une nouvelle partition du lac de données ? Vous pouvez éviter les tâches fastidieuses et lentes en ajoutant simplement un nouveau magasin de données comme destination pour vos événements existants.

Posséder son infrastructure d'analyse signifie donner à l'équipe chargée des données le contrôle total et la flexibilité nécessaire pour s'adapter rapidement aux besoins de l'entreprise sans être confronté aux limites douloureuses de la dépendance vis-à-vis d'un fournisseur.

Enregistrer chaque visite de site

L'enregistrement de toutes vos données nécessite l'utilisation d'un outil d'analyse spécialement conçu pour collecter des données brutes et de première main. Les SDK de RudderStack sont spécialement conçus pour enregistrer les événements des clients et peuvent être mandatés derrière l'URL de votre site ou de votre application, ce qui vous permet d'enregistrer l'intégralité des données.

Visits captured



Une vue d'ensemble avec des charges utiles riches, configurables et transformables

Les charges utiles prêtes à l'emploi de RudderStack fournissent des données beaucoup plus riches que celles de GA4. Elles présentent également un schéma JSON normalisé et ouvert, avec des objets que vous pouvez personnaliser avec des propriétés d'événement et des caractéristiques d'utilisateur qui répondent aux besoins spécifiques de votre entreprise. Si les besoins d'analyse changent, ce qui est toujours le cas, vous pouvez utiliser les transformations de RudderStack pour renommer les clés ou remodeler les charges utiles sans avoir à toucher au code dans votre application ou sur votre site web.

Charge utile de la page GA4	Charge utile de la page RudderStack
-----------------------------	-------------------------------------

```
v: 2
tid: G-252450485
gtm: 2oebu0
_p: 309274267
cid: 679471488.1651608029
ul: en-us
sr: 2560x1440
uaa:
arm
uab: 64
uafvl:
Google%20Chrome;107.0.5304.121|Chromium;1
07.0.5304.121|Not%3DA%3FBrand;24.0.0.0
uamb: 0
uam:
uap: macOS
uapv: 12.6.0
uaw: 0
_s: 2
dl: /pricing/
dr: $direct
dt: RudderStack Pricing
sid: 1671219066
sct: 1
seg: 1
en: page_view
_ee: 1
```

```
{
  "channel": "web",
  "context": {
    "app": {
      "build": "1.0.0",
      "name": "RudderLabs JavaScript
SDK",
      "namespace":
"com.rudderlabs.javascript",
      "version": "1.21.0"
    },
    "traits": {
      "name": "First Name"
    },
    "library": {
      "name": "RudderLabs JavaScript
SDK",
      "version": "1.21.0"
    },
    "userAgent": "Mozilla/5.0
(Macintosh; Intel Mac OS X 10_15_7)
AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/107.0.0.0 Safari/537.36",
    "locale": "en-US",
    "os": {
      "name": "",
      "version": ""
    },
    "screen": {
      "density": 2,
      "width": 2560,
      "height": 1440,
      "innerWidth": 1280,
      "innerHeight": 787
    },
    "sessionId": 1671218225671,
    "campaign": {},
    "page": {
      "path": "/pricing/",
      "referrer": "$direct",
      "referring_domain": "",
      "search": ""
    }
  }
}
```

```
        "title": "RudderStack pricing",
        "url":
"https://rudderstack.com/pricing/",
        "tab_url":
"https://rudderstack.com/pricing/",
        "initial_referrer": "$direct",
        "initial_referring_domain": ""
    }
},
"type": "page",
"messageId":
"1671218575125100-c1271cc4-5b0c-4b99-92a7-58
cb69310802",
"originalTimestamp":
"2022-12-16T19:22:55.125Z",
"anonymousId":
"8bd1d069-0318-4d97-bfca-6b4a39ab0874",
"userId": "eric@rudderstack.com",
"properties": {
    "path": "/pricing/",
    "referrer": "$direct",
    "referring_domain": "",
    "search": "",
    "title": "RudderStack Pricing",
    "url":
"https://rudderstack.com/pricing/",
    "tab_url":
"https://rudderstack.com/pricing/",
    "initial_referrer": "$direct",
    "initial_referring_domain": ""
},
"integrations": {
    "All": true
},
"sentAt": "2022-12-16T19:22:55.135Z"
}
```

Instaurez la confiance grâce à la transparence totale de votre entrepôt

Lorsque vous stockez et modélisez toutes vos données dans votre entrepôt, vous n'avez jamais à vous demander ce que fait GA4 à l'intérieur de la boîte noire et si ses algorithmes font le bon choix pour votre entreprise.

« Nous utilisons la version gratuite de Google Analytics. Tout était anonymisé et nous ne pouvions pas voir ce que nos utilisateurs faisaient sur notre site web ou comment ils utilisaient notre produit. RudderStack nous a permis d'améliorer notre visibilité en ce qui concerne le comportement et le parcours des utilisateurs, de mieux comprendre notre entonnoir et de réaliser des tests A/B qui nous permettent de personnaliser l'expérience utilisateur. »

Mona Sami, Directrice de l'analyse des données, InfluxData

```
1 select * from "RUDDER_WEBAPP_DATA"."RUDDER_WEBAPP"."IDENTIFIES"
2
3
```

Results Data Preview

✓ Query ID SQL 789ms 100 rows

Filter result...  Copy

Row	NAME	ANONYMOUS_ID	ID	CHANNEL	CONTEXT_IP
1	Steve Jobs	5d2fd9c5-5231-4b36-96da-...	dd5288ec-667a-4b78-9590-cef69...	web	96.232.85.15
2	Elon Musk	0eb93ac7-9046-404a-a2ce-...	d7a91de0-d510-44f8-bc4f-93b60a...	web	170.10.252.83
3	Walt Disney	81e84cf3-5848-4836-86d9-...	7d11407a-bef6-4516-80d3-f7eb72b...	web	134.41.188.254

Faire des analyses sur votre entrepôt vous offre une transparence et une flexibilité totales.

Déployer et modifier des modèles de données flexibles qui correspondent au comportement réel

En utilisant des outils comme dbt, vous pouvez élaborer vos propres modèles en ce qui concerne les cas d'utilisation principaux que sont la résolution d'identité et la sessionnalisation, puis les modifier au fur et à mesure que vous apprenez comment vos utilisateurs interagissent avec votre site web et vos applications.

Se libérer du chaos de Google en matière de conformité

Plus important encore, le fait d'être propriétaire de votre infrastructure d'analyse signifie que vous pouvez dire adieu au chaos de Google en matière de conformité une fois pour toutes, ce qui signifie que vos équipes chargées de la sécurité informatique, de la législation et du marketing peuvent dormir en toute tranquillité.

Guide pratique : Démarrer avec l'analyse de l'entrepôt en premier lieu

Dans cette section du guide, nous aborderons la définition de l'analyse de l'entrepôt en premier lieu et nous passerons en revue les étapes du remplacement de GA4 par votre propre infrastructure de données.

Qu'est-ce que l'analyse de l'entrepôt en premier lieu ?

Une pile d'analyse de données de l'entrepôt en premier lieu est une pile d'analyse avec un seul référentiel de données (également connu sous le nom d'entrepôt de données) pour toutes les données des clients. Comme nous l'avons mentionné plus haut, les outils d'analyse tels que GA4 créent des silos de données, mais il en va de même pour d'autres référentiels de données clients tels que Salesforce. Il est donc difficile pour les entreprises d'accéder à toutes les données dont elles ont besoin pour prendre des décisions complexes.

En rassemblant toutes les données de votre entreprise en un seul endroit, y compris les données de trafic de vos sites Web et applications, vous êtes en mesure d'élaborer des analyses efficaces sur des ensembles de données importants, diversifiés et de haute qualité pour répondre à des questions auxquelles des outils d'analyse tels que GA4 ne sont pas en mesure de répondre seuls.

Supposons par exemple que votre entreprise soit une startup en phase de démarrage. Dans ce cas, vous essayez probablement de comprendre comment les clients perçoivent votre produit ou interagissent avec lui. Ainsi, vous avez certainement des questions du genre :

- Combien d'utilisateurs actifs avons-nous eu ce mois-ci ?
- Quel est notre taux de rétention global ?
- Quelles sont nos principales sources d'inscription ?

Pour ces questions de base, GA4 peut faire le travail (bien qu'avec des données limitées et cloisonnées), mais à mesure que votre entreprise se développe, vous aurez besoin de réponses à des questions plus complexes telles que :

- Quel comportement au cours de la première journée d'un utilisateur indique qu'il est susceptible de souscrire à un plan payant ?
- Comment la valeur de la durée de vie varie-t-elle en fonction du canal publicitaire payant ?

- Quels modèles conduisent à une augmentation des achats de nos produits ?
- Quels produits devrions-nous recommander à des utilisateurs spécifiques pour augmenter les revenus ?
- Comment devons-nous fixer le prix de notre produit en fonction du coût d'acquisition ou du coût des ressources par client ?

Vous ne pouvez pas répondre à ces questions avec GA4. Mais vous le pouvez, avec l'analyse de l'entrepôt en premier lieu. Si vous disposez de toutes les données dans votre entrepôt, vos équipes en charge des données et de l'analyse peuvent non seulement modéliser les données brutes pour trouver les réponses, mais aussi comprendre comment ces réponses évoluent dans le temps, le tout avec des requêtes totalement transparentes et contrôlables par version.

Les composants d'une pile d'analyse de données axée sur l'entrepôt en premier lieu

Une pile d'analyse de données d'entrepôt en premier lieu est constituée de quelques composants essentiels pour faciliter la collecte, le stockage et le déplacement de vos données afin d'en faciliter l'analyse :

1. Entrepôt de données dans le cloud
2. Ingestion de données comportementales
3. Modélisation des données
4. Visualisation

Pour chaque élément, examinons quelques outils adaptés à la tâche à accomplir.

1. Centraliser les données : votre entrepôt de données

Grâce aux données brutes de trafic et de comportement, ainsi qu'à toutes les autres données de vos clients, réunies en un seul endroit, vous êtes totalement propriétaire de vos données et vous pouvez effectuer des analyses beaucoup plus approfondies. Cela signifie également que vous pouvez facilement passer d'un outil d'analyse à l'autre et l'intégrer à différentes équipes, car chaque outil a accès à l'ensemble de vos données historiques.

De nombreuses options d'entrepôt de données sont disponibles, mais Redshift, Snowflake et BigQuery sont les plus adoptées. Bien que chacun de ces entrepôts ait ses propres avantages, ils fonctionnent tous bien pour intégrer des données utiles d'événements à partir d'un outil comme RudderStack afin d'effectuer des analyses web. Les entrepôts ne servent pas qu'aux analyses web, et les équipes chargées des données devront donc se pencher sur d'autres questions, telles que les autres types de données stockées (c'est-à-dire les données semi-structurées), le niveau de gestion des performances requis par les ingénieurs et l'évolution des coûts en fonction du cas d'utilisation de votre entreprise.

2. Collecte des données : ingestion d'événements

Une fois que vous avez un lieu de centralisation de vos données, vous devez commencer à les collecter. L'ingestion est un sujet important lorsqu'il s'agit d'introduire des données dans votre entrepôt. Dans ce guide, nous allons donc nous concentrer sur ce dont vous avez besoin pour remplacer GA4 dans le cadre de l'analyse web et mobile.

Pour élaborer des analyses web et mobiles, vous devez enregistrer des données sur le comportement des utilisateurs, qui se présentent sous la forme d'événements et de caractéristiques de l'utilisateur. Les événements décrivent une action de l'utilisateur, comme un achat, tandis que les traits décrivent l'utilisateur individuel qui a effectué cette action.

Comme les équipes modernes en charge des données le savent si bien, Google Analytics est un silo notoire pour ce type de données comportementales et, même si vous pouvez techniquement exporter les données par le biais de tâches ETL ou, avec GA4 vers les produits Google, elles nécessitent toujours une tonne de modélisation pour être utilisables, et cela avant même d'essayer de comprendre comment joindre les identités des utilisateurs à d'autres types de données des clients provenant d'autres sources.

Tous ces problèmes sont aggravés lorsque des scripts supplémentaires provenant d'autres outils d'analyse sont ajoutés à l'équation, créant ainsi de multiples pipelines d'ingestion dédiés qui alimentent leurs propres silos.

Une source unique de vérité pour l'ingestion des données

Les équipes chargées des données qui élaborent des analyses sur leur entrepôt utilisent un outil spécialisé comme RudderStack pour l'ingestion des données comportementales. Avec RudderStack, vous pouvez instrumenter votre site avec un seul SDK, enregistrer les événements et les charges utiles d'identification de l'utilisateur en une seule fois, puis syndiquer cette version

unique de la vérité à la fois à votre entrepôt et à d'autres outils d'analyse SaaS en aval (plus d'informations sur les intégrations SaaS ci-dessous).

Les équipes chargées des données se réjouissent : un SDK unique signifie qu'il n'y a plus de scripts gonflés, de séquençage ou de gestion des changements d'API et des pipelines interrompus. L'approche de RudderStack est également utile lorsqu'une équipe souhaite changer d'outil et ajouter un nouvel outil à la pile. Vous n'aurez pas besoin de toucher à la base de code pour supprimer ou ajouter de nouveaux SDK. En revanche, vous pouvez simplement supprimer l'outil ou intégrer le nouvel outil via le tableau de bord de l'outil de diffusion d'événements.

Lorsqu'il s'agit d'analyser votre entrepôt, les données utiles de RudderStack sont automatiquement aplaties et chargées dans des tableaux spécifiques selon un schéma standardisé, ce qui signifie que votre équipe d'analyse a déjà une longueur d'avance lorsqu'il s'agit de créer des jointures et de modéliser des données.

Voici quelques caractéristiques spécifiques du pipeline de diffusion d'événements de RudderStack que nos utilisateurs trouvent particulièrement utiles pour construire des analyses comportementales riches :

- Données de suivi de session configurables dans chaque charge utile
- Propriétés des événements et caractéristiques des utilisateurs configurables
- Résilience des bloqueurs de publicité
- Suivi automatique des utilisateurs anonymes à l'aide d'un identifiant unique
- Plusieurs méthodes d'identification des utilisateurs et d'attribution des caractéristiques aux utilisateurs et aux groupes, même entre domaines

3. Préparer les données pour l'analyse : la modélisation des données

Une fois que votre pipeline d'ingestion de données comportementales a récupéré les événements de vos sites Web et applications, l'étape suivante consiste à donner un sens à toutes ces données.

La modélisation des données consiste à organiser, transformer et regrouper les événements pour répondre aux questions auxquelles les équipes ont formellement tenté de répondre avec GA4. Par exemple, supposons que votre équipe marketing souhaite comprendre quels sont les canaux qui génèrent le plus grand nombre de consultations de contenu (c'est-à-dire la recherche

organique par rapport à la recherche sociale payante). Vous allez devoir compter les pages vues pour chaque URL concernée, puis les regrouper par domaine de référence.

Il convient de noter que les schémas d'événements de RudderStack sont spécifiquement conçus pour faciliter ce type de modélisation. Pour effectuer l'analyse ci-dessus, vous devez compter les pages vues par différent « anonymousId » (qui représente un utilisateur unique, anonyme ou connu) pour obtenir un nombre unique de pages vues, puis utiliser la valeur de « référence » pour regrouper les pages vues en fonction du site qui les a référencées. Pour maintenir la continuité avec le regroupement des canaux par défaut de Google Analytics, vous devez attribuer un nom de canal à un groupe d'URL (par exemple, Twitter, LinkedIn et Facebook peuvent être regroupés dans un canal appelé « Social »).

Une fois les données modélisées, il est plus facile d'en extraire de la valeur, que ce soit sous forme de tableaux de bord, de rapports ou de base pour l'analyse prédictive ou prescriptive.

De nombreuses équipes en charge des données utilisent SQL ou Python pour modéliser leurs données, mais ces dernières années, des outils comme dbt ont significativement amélioré le flux de travail de modélisation pour les équipes en charge des données.

Qu'est-ce que dbt ?

L'un des outils de modélisation de données les plus populaires est [dbt](#) conçu par [dbtlabs](#). dbt (outil d'élaboration des données) permet aux ingénieurs en analyse de données de transformer les données de leurs entrepôts en écrivant simplement des instructions de sélection SQL. dbt prend ces codes SQL et les exécute dans votre entrepôt de données pour créer des tableaux et des vues.

dbt permet aux ingénieurs en données de travailler comme des ingénieurs en logiciel avec le contrôle de version, l'intégration continue et les tests intégrés. Comme dbt est basé sur SQL, il est facile à utiliser. De plus, dbt est open-source et dispose d'une communauté très active sur Slack.

RudderStack propose plusieurs modèles dbt qui, lorsqu'ils sont exécutés sur des données d'événements provenant de RudderStack, peuvent constituer des modèles de données presque complets pour les principaux cas d'utilisation de l'analyse Web :

- sessionalisation ;
- cartographie du parcours client ;
- résolution de l'identité ;

- analyse web.

Consultez [l'article sur la base des connaissances](#) pour en savoir plus sur les modèles dbt et sur la manière de les utiliser pour élaborer des analyses.

4. Exploitation des données : outils de visualisation

La modélisation des données n'est pas complète sans visualisation. Maintenant que vous avez transformé les données de l'entrepôt en ensembles de données riches, l'étape suivante consiste à les introduire dans un outil de visualisation qui vous permettra de représenter visuellement ces ensembles de données sous forme de diagrammes, de cartes, de graphiques ou d'images afin d'en tirer des informations précieuses.

Il existe sur le marché de nombreux outils de visualisation axés sur les entrepôts. Parmi les plus populaires, se trouvent : Tableau, Metabase, Sigma et Hex. Toutefois, avant de choisir un outil, vous devez prendre en compte des éléments tels que la facilité d'utilisation, la courbe d'apprentissage, la prise en charge de l'entrepôt de données de votre choix, le niveau de flexibilité, les options de personnalisation et l'adéquation de l'outil à votre cas d'utilisation.

Tableau

[Tableau](#) est souvent reconnu comme le roi incontesté des logiciels de visualisation de données, et ce à juste titre, en raison de sa facilité d'utilisation et de sa capacité à générer des visuels interactifs bien au-delà des solutions BI standard. Il est particulièrement bien adapté au traitement d'ensembles de données volumineux et en évolution rapide.

Metabase

[Metabase](#) est un outil de veille stratégique open-source qui vous permet de créer des graphiques et des tableaux de bord en utilisant des ensembles de données provenant de votre entrepôt de données. Bien que SQL ne soit pas nécessaire pour produire des visualisations, Metabase autorise SQL à procéder à une personnalisation renforcée. Sa simplicité et sa facilité d'utilisation sont les principales raisons pour lesquelles de nombreux utilisateurs l'apprécient.

Sigma

[Sigma](#) est une plateforme d'analyse cloud qui utilise une interface de feuille de calcul familière pour donner aux utilisateurs professionnels un accès instantané à l'exploration et à la compréhension de leur entrepôt de données cloud.

Hex

[Hex](#) est une plateforme moderne pour la science et l'analyse des données. Des analyses ad-hoc aux magnifiques tableaux de bord en passant par des applications complexes, Hex rationalise l'ensemble du flux de travail analytique, de sorte que votre équipe puisse se concentrer sur la génération d'informations, la prise de décisions et l'avancement des activités.

Des données riches et cohérentes pour l'ensemble de votre pile

RudderStack fournit la couche de données qui vous permet de tirer le meilleur parti de GA4, de débloquent des analyses plus approfondies dans votre entrepôt de données et d'envoyer des données à tous les autres outils de votre pile. Notre grande bibliothèque d'intégration comprend plus de 180 outils utilisés au quotidien par les équipes en charge du marketing, de la vente, des produits et du suivi de la clientèle. Inscrivez-vous gratuitement pour commencer à améliorer vos analyses dès aujourd'hui.

Le guide

Comment franchir les limites d'AG4 ?

Table des matières

01. Avantages du suivi côté serveur

02. Options du suivi côté serveur de Google Analytics 4

03. Architecture d'un déploiement hybride

04. Utilisation du mode hybride de RudderStack

05. Aller au-delà de GA4

Collecte de données par des tiers

La performance du site en prend un coup

Sensible aux bloqueurs de publicité

Prise en charge des principales fonctionnalités

Collecte de données de première partie

Performances optimales du site

Bloqueur de publicité résilient

Ne prend pas en charge les principales fonctionnalités

Choisissez votre mode de connexion

SDK Web

Mode hybride



 rudderstack.com

 [@rudderstack](https://twitter.com/rudderstack)