Segment warehouse enabled before you enable these. From the warehouse, you can analyze your data with SQL, use Reverse ETL to extract data, or use Engage SQL Traits to build audiences. Some examples of Object Cloud sources are Salesforce (account information), Zendesk (support cases), and Stripe (payments information).

> ℹ️
> In the app, data from website, mobile, and server sources can go to a warehouse **or** to destinations. Object Cloud-App Source data can **only** go to Warehouses.

How do cloud sources work?

Sources are functionally comprised of either one or both of the following components: a "sync" component and a "streaming" component. They work together to populate logical collections of data based on upstream resource availability and following data normalization best practices. These collections may be either events (append only data streams, akin to "facts" in data warehousing parlance) or objects (dimensional values that may be updated based on changes in state upstream).

Sync frequency

You enable a cloud source from the Segment web app, and grant Segment access by pasting an API key or authenticating with OAuth. Segment then starts a scheduled job on your behalf which makes requests to the downstream tool, normalizes and transforms the data, and forwards that data to the Segment API.

Cloud sources attempt to use as few API calls as possible, and (where possible) only fetch data that changed since the last sync. The syncs might take a long time (especially on the first sync), so the cloud source syncs have robust retry and rate limiting logic.

Contact Segment Product Support if you'd like to change the cadence of your source's sync frequency.

API call use and collection selection

We make an effort to be respectful of your API call allotments and limits. For example, in the case of Salesforce, we issue only one query per collection per run, using the absolute minimum number of API calls possible (typically about 350/day).

Moreover, we're deliberate about which collections we pull, striking a balance between allowing you to get a full picture of your users and reducing extraneous data (like administrative and metadata tables).

Soon, we'll allow you to specify which collections you care about during the source set up phase, so if you need to cut down on calls, you'll be able to just deselect collections.

Streaming

Streaming components are used to listen in real time to webhooks from downstream cloud sources, normalize and transform the data, and forward it to our APIs.

Both sync and streaming components can forward data to our event tracking and objects upsertion API processing layers, but generally sync components are used to fetch objects and streaming components listen for events.

Set up a cloud source

To use cloud sources, we suggest going through the following steps.

1. Get cloud source credentials
2. Get warehouse credentials
3. Choose your preferred sync time

Before you connect a source, check out the sources documentation. See what kind of credentials you will need to enable the source. Different sources require different levels of permissioning.

Next, you'll also need to get the credentials for your warehouse.

Once you have the necessary credentials (or are logged in to OAuth for your cloud source), you should be ready to go!

1. Go to the "sources catalog" in the Segment web app.
2. Choose a cloud source, and click Configure.
3. Enter your credentials or log in using OAuth.
4. Go to the "warehouses" tab and enter the credentials for your warehouse if you don't already have one connected to Segment.

Based on your plan, you can schedule a certain number of syncs per day. We suggest setting these up so your dashboards and reports are fresh for reporting, but not at the same time of day that a lot of people are querying your database.

Troubleshooting cloud sources

The most common reason cloud sources have trouble because of authentication or permission issues. When the issue is related to authentication, you'll see an "access denied" connection error in your source details. When this happens, Segment quits the process early and does not make any further attempts on any collections.

When you successfully authenticate, but your user lacks the required permissions (for example, if you use an agent login instead of an administrator for Zendesk), Segment attempts to pull each collection and reports errors on a per-collection basis. This helps you troubleshoot why source runs fail, because sometimes permission-based denials are scoped to specific resources from the upstream tool.

Segment attempts to make the errors displayed in the UI clear enough so we don't need to document all of them. However, if it's not clear what to do to fix an error you encounter, contact support and let them know.

Sometimes, when the sync job fails due to an unhandled error or is mysteriously hanging for too long, we'll kill the job and report a failure with instructions to contact support. When this happens, our support and engineering teams have already been notified of the failure and have the complete set of logs to set about debugging and remediating the issue, but don't hesitate to get in touch so they can keep you in the loop!

Using Cloud Source data

What kind of data does Segment pull from each source?

In general, we've focused on pulling all of the collections directly related to the customer experience. We do not automatically pull all collections available from a partner API, since many of them aren't relevant to the customer journey. You can see a list of the collections we pull in the docs for each cloud source. Each collection reflects a table in your database.

Contact Segment Product Support if you need additional data collected, or to change the schema to do the analysis you want. We'd love to know what analysis you're trying to run, what additional data you need, and we'll share with the product team to evaluate.

What questions can you answer with data from cloud, web, and mobile sources combined in a single warehouse?

- What content drives people forward in our sales funnel?
- What are the top pages viewed before a support ticket is sent?
- Do people who opt into text messages engage more than people who only get emails?
- Do customers that interact with our support team activate faster? - Retain more overtime?
- What are all of the communications across marketing, success, and sales, this account has had in the last 2 months?

Querying source data

Generally, you need intermediate- to advanced SQL experience to explore and analyze cloud source data in a warehouse. The following resources can help you get up and running more quickly!

Joining IDs As you start to get into joining across different types of sources, you'll need a way to join user IDs. This help article explains how to do this in detail.

Partner Dashboards Our BI partners at Mode, Looker, BIME, Periscope, and Chartio have created out of the box dashboards that work on top of our source schemas.

This page was last modified: 22 Oct 2024

Need support?

Questions? Problems? Need more info? Contact Segment Support for assistance!

Visit our Support page

Help improve these docs!

📝 Edit this page

⊕ Request docs change

Was this page helpful?

👍 Yes

👎 No

Get started with Segment

Segment is the easiest way to integrate your websites & mobile apps data to over 300 analytics and growth tools.

Your work e-mail

Request Demo

or

Create free account