Gatsby contains a telemetry feature that collects anonymous usage information that is used to help improve Gatsby for all users. The Gatsby user base is growing very rapidly. It's important that our small team and the greater community will better understand the usage patterns, so we can best decide how to design future features and prioritize current work.

You will be notified when installing Gatsby and when running it for the first time.

How to opt-out

Users may always opt out from the telemetry with `gatsby telemetry --disable` or setting the environment variable `GATSBY_TELEMETRY_DISABLED` to `1`

Why?

Anonymous aggregate user analytics allow us to prioritize fixes and features based on how and when people use Gatsby. Since much of Gatsby's function revolves around community plugins and starters, we want to collect information on usage and reliability so that we can ensure a high-quality ecosystem.

This raises a question: how will we use this telemetry data to improve the ecosystem? Some examples are helpful:

- We will be able to understand which plugins are typically used together. This will enable us to surface this information in our public plugin library and build more relevant starters and tutorials based on this data.
- We will be able to surface popularity of different starters in the starter showcase.
- We will be able to get more detail on the types of errors users are running into in *every* build stage (e.g. development, build, etc.). This will let us improve the quality of our tool and better focus our time on solving more common, frustrating issues.
- We will be able to surface reliability of different plugins and starters, and detect which of these tend to error more frequently. We can use this data to surface quality metrics and improve the quality of our plugins and starters.
- We will be able to see timings for different build stages to guide us in where we should focus optimization work.

What do we track?

We track general usage details, including command invocation, build process status updates, performance measurements, and errors. We use these metrics to better understand the usage patterns. These metrics will directly allow us to better decide how to design future features and prioritize current work.

Specifically, we collect the following information for *all* telemetry events:

- Timestamp of the occurrence.
- Command invoked (e.g. `build` or `develop`).
- Gatsby machine ID: This is generated with UUID and stored in global gatsby config at `~/.config/gatsby/config.json` .
- Unique session ID: This is generated on each run with UUID.
- One-way hash of the current working directory or a hash of the git remote.
- General OS level information (operating system, version, CPU architecture, and whether the command is run inside a CI).
- Current Gatsby version.

The access to the raw data is highly controlled, and we cannot identify individual users from the dataset. It is anonymized and untraceable back to the user.

What about sensitive data? (e.g. secrets)

We perform additional steps to ensure that secure data (e.g. environment variables used to store secrets for the build process) **do not** make their way into our analytics. [We strip logs, error messages, etc.](#) of this sensitive data to ensure we *never* gain access to this sensitive data.

You can view all the information that is sent by Gatsby's telemetry by setting the environment variable `GATSBY_TELEMETRY_DEBUG` to `1` to print the telemetry data instead of sending it over.