Technical white paper

December 2025

How to connect GA4Dataform and GrowthBook to reduce BigQuery costs by 82%

GA4Dataform
by Superform Labs

A case by Superform Labs, Mintminds and The Social Hub

Table of contents

Who are we? **3**
- Superform Labs 3
- Mintminds 3
- The Social Hub 3

Business case **4**
- Management Summary 4
- The challenge 4
- The problem: querying the inefficient GA4 tables 4
- The solution: use GA4Dataform flattened tables 4
- Technical features of the solution 5
- The proof: an experiment 5
- The conclusion 6
- Full details 6

Setup Guide **7**
- Installing GA4Dataform 7
- Configuring your custom dimensions 9
- Run the model 10

Setting up the GrowthBook queries **11**
- Assignment query 11
- Fact table queries 12
- Metrics, Dimensions and Segments queries 13

Thanks to… **14**

Links and resources **15**
- GrowthBook 15
- Mintminds 15
- GA4Dataform by Superform Labs 15
- The Social Hub 15

Who are we?

Superform Labs 🌐

Unlock the power of your data. We provide pre-built Dataform models for BigQuery, making your Google Analytics and Google Ads data instantly usable and fully orchestrated inside your Google Cloud Project.

Mintminds 🌐

Mintminds builds enterprise-grade experimentation programs for ambitious brands. We specialize in advanced GrowthBook infrastructure—enabling teams to run more experiments, faster, with reliable data—scaling from 10 to 100+ experiments per month at a fraction of traditional vendor costs.

The Social Hub 🌐

More than just a place to stay. We are a hybrid hospitality brand offering hotel rooms, co-living spaces, coworking spaces, and creative playgrounds across Europe.

Business case

Management Summary

In this business case, we have demonstrated how to eliminate the unpredictable variable cost of BigQuery in large-scale experimentation. Using two setups and comparing the results, we saw that total BigQuery costs were reduced by 82%.

The challenge

The core challenge for companies running active experimentation programs (like those using **GrowthBook with GA4 as a data source**) is the high and unpredictable variable cost of BigQuery.

The problem: querying the inefficient GA4 tables

GA4's standard and intraday tables store event parameters in nested structures, forcing BigQuery to scan and process significantly more data than is necessary, even when only a few metrics and dimensions are needed for an experiment.

Unoptimized BigQuery data processing can easily reach hundreds of Gigabytes per experiment query.

The solution: use GA4Dataform flattened tables

By connecting GrowthBook to the output tables of **GA4Dataform** by Superform Labs, we can use a much more efficient, flattened dataset optimized for the type of queries GrowthBook uses.

Technical features of the solution

Feature	Benefit
Fully Flattened Structure	Eliminates nested fields, leading to dramatically faster queries.
Smart Partitioning & Clustering	Tables are optimized for the most commonly used filters.
Smart queries leveraging the partitions and clusters	Decreases the number of rows scanned by restricting queries by date and event name.
Daily Incremental Updates	Ensures fresh data is available using efficient logic.

This optimized dataset is significantly cheaper to query than repeatedly querying GA4's nested tables. It also compounds savings by being usable for all other BigQuery reports and dashboards.

The proof: an experiment

Mintminds partnered with The Social Hub to conduct an objective, week-long A/A test. Two identical experiments were run simultaneously:

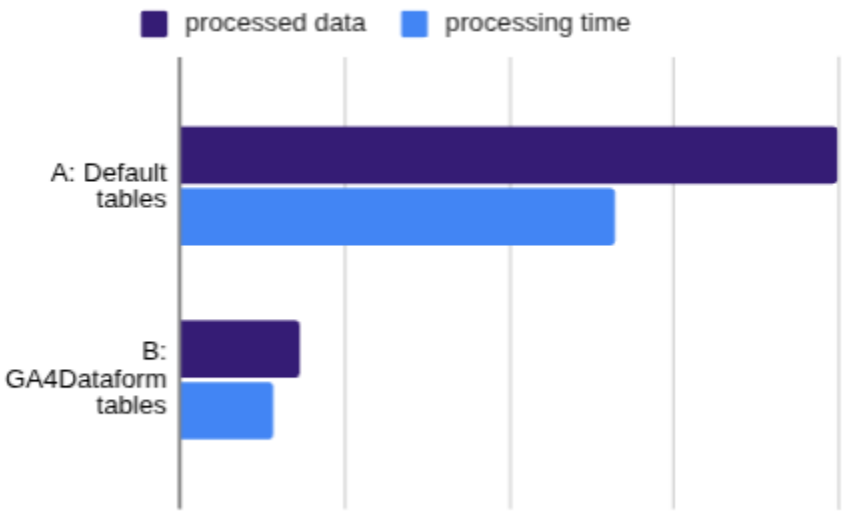
- **Experiment A**: Used Default GrowthBook queries (nested GA4 tables).
- **Experiment B**: Used Optimised queries (flattened GA4Dataform dataset)

We used BigQuery query cost logs to analyse the cost and speed of the used queries to run and analyse these tests.

The results showed a **massive cost reduction** by switching to the GA4Dataform flattened dataset:

- 81.8% Query Cost Reduction
- 77.5% Faster Query Speed (from over a minute to just seconds)

This reduction ensures better monitoring capabilities without budget concerns and allows for a significant scale-up in experiment volume.



The conclusion

By combining **GrowthBook Pro**, **GA4Dataform Core**, and strategic BigQuery optimization, businesses can build a cost-effective, high-performance experimentation stack that rivals Enterprise setups at a fraction of the price.

Full details

Mintminds published full details of the case study on their website (link)

Setup Guide

In the following chapters, we will describe the full details on how to set up GA4Dataform and how to use GrowthBook to use the optimized datasets.

- install GA4Dataform (10-20 minutes)
- complete the GA4Dataform setup (10 minutes)
- configure the GrowthBook queries (10 minutes)

The complete setup will take you less than an hour

Note: The full GA4Dataform guide can be found at docs.ga4dataform.com.

⇛ This guide will assume you have access to the Google Cloud Project where your GA4 export is also located.

The full GCP permission requirements can be found here

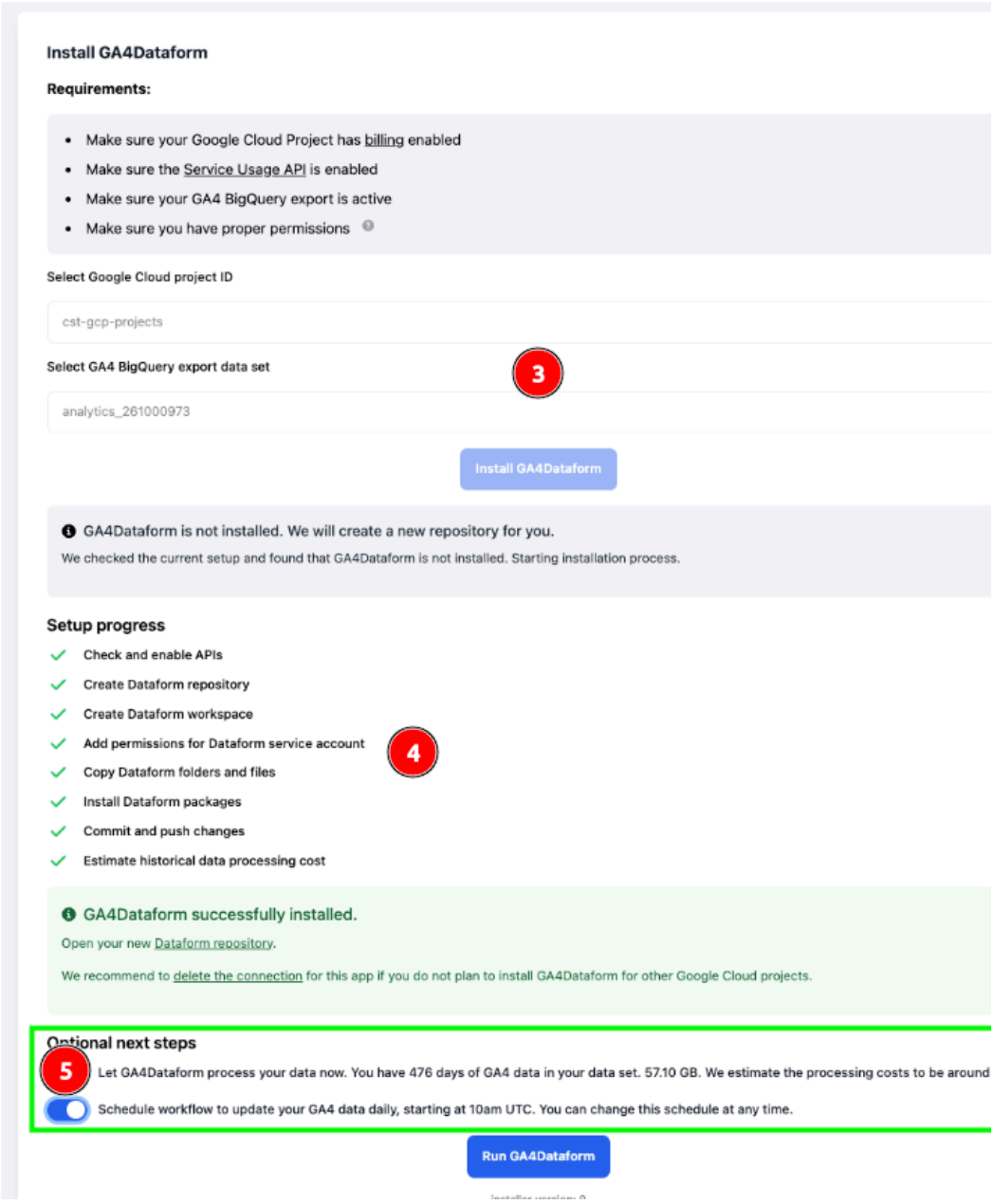
Installing GA4Dataform

GA4Dataform has both open source and commercial licenses. For this walkthrough, we will use GA4Dataform Core product, which installs via the one-click installer and comes at zero cost.

1. Obtain a license key for the software at the GA4Dataform website
   Choose "Core" and a free license key will be mailed to you
2. Save the license key, and click "Proceed to GA4Dataform installation"
3. In the installer, enter your license key, select the GCP project and your analytics dataset and click "Install"
4. The install process will run for a few minutes

5. At the end of the installation process, choose **NOT** to let GA4Dataform process historical data - we will do that manually.

After the install, you should see the end screen, and click "Open your Dataform repository" in the green box.



Configuring your custom dimensions

There is a lot to customize if you want (see the full documentation here), but for now, we will focus on adding the custom GA4 parameters that you configured for GrowthBook to use.

Usually, these parameters are:

- **`exp_id`** - for the experiment ID
- **`exp_var`** - for the variation ID

In your GA4Dataform repository, add these parameters in the **`CUSTOM_EVENTS_PARAMS_ARRAY`** section of **`includes/custom/config.js`**

```
CUSTOM_EVENT_PARAMS_ARRAY: [

  { name: "exp_id", type: "string", description: "GrowthBook experiment ID" },
  { name: "exp_var", type: "int", description: "GrowthBook experiment variation" }

],
```

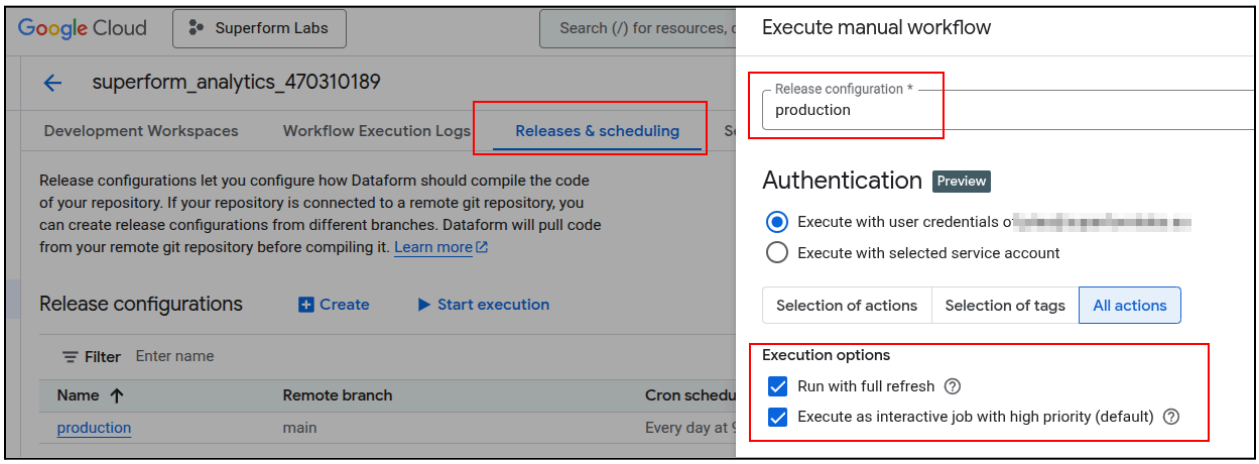
Save your changes to production, by clicking "COMMIT" in the top left corner, followed by "PUSH".

Run the model

GA4Dataform needs to process your GA4 data at least once, to do a backfill. This is called a "Full refresh".

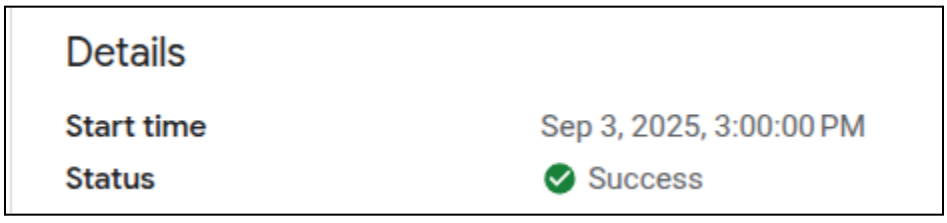
In the daily scheduled run (if configured), the model will use incremental updates (to save data processing). See the documentation section for details.

Triggering a full refresh, to build the table with the freshly configured dimensions included can be done in the "Releases and scheduling" tab.



You can follow the progress of the run in the "Workflow Execution Logs" section.

If everything is done, you should see the green checkmark.



Setting up the GrowthBook queries

When the flat events table is generated by GA4Dataform, we can use it in GrowthBook.

The output table can be found here:
`` `project.superform_outputs_NNNN.ga4_events` ``

Where NNNN is the same number that GA4 uses in the name of your raw table dataset: `` `project.analytics_NNNN.events_*` ``

⇛ The `ga4_events` table is partitioned by the column **`event_date`**, and clustered by the column **`event_name`**.

Try to always use these columns in the WHERE clause - so you save data processing cost

Assignment query

Under "Data Sources" in GrowthBook, add the following Assignment Query:

Note: in GA4, we configured the event name for GrowthBook to be `` `gb_view_experiment` ``.

```
SELECT
  user_pseudo_id AS anonymous_id,
  time.event_timestamp_utc AS timestamp,
```

```
  event_params_custom.exp_id AS experiment_id,
  event_params_custom.exp_var AS variation_id,

  -- optional: add dimension columns
  device.category AS device_category,
  device.web_info.browser,
  device.operating_system


FROM `projectname.superform_outputs_8008135.ga4_events`

--- filter on event name and date
WHERE event_name = 'gb_view_experiment'
  AND event_date BETWEEN
     '{{date startDateISO "yyyy-MM-dd"}}' AND '{{date endDateISO "yyyy-MM-dd"}}'

-- exclude hits without anonymous ID
  AND user_pseudo_id IS NOT NULL
```

This query uses the custom dimensions **`exp_id`** and **`exp_var`** we created in GA4Dataform. These live inside the struct **`event_params_custom`**.

Fact table queries

To construct a fact table query, the same flat table can be queried. Make sure to use data-saving columns for date and event name.

Example, for a Purchase face table with revenue, quantity and coupon:

```
SELECT
  user_id,
  user_pseudo_id AS anonymous_id,
  time.event_timestamp_utc AS timestamp,
  ecommerce.purchase_revenue,
  ecommerce.total_item_quantity,
  event_params.coupon,
FROM `projectname.superform_outputs_8008135.ga4_events`

--- filter on event name and date
WHERE event_name = 'purchase'
```

```
  AND event_date BETWEEN
     '{{date startDateISO "yyyy-MM-dd"}}' AND '{{date endDateISO "yyyy-MM-dd"}}'

-- exclude hits without anonymous ID
  AND user_pseudo_id IS NOT NULL
```

The `coupon` field is one of the many standard event parameters in the flat table.

Metrics, Dimensions and Segments queries

If you don't use face table, but want to construct metrics, dimensions, and segments queries in GrowthBook: this follows the same pattern, here's an example of a metrics query named **Login Page Viewed**:

```
SELECT
  user_id,
  user_pseudo_id AS anonymous_id,
  time.event_timestamp_utc AS timestamp,
  1 as value
FROM `projectname.superform_outputs_8008135.ga4_events`

--- filter on event name and date
WHERE event_name = 'page_view'
  AND event_date BETWEEN
     '{{date startDateISO "yyyy-MM-dd"}}' AND '{{date endDateISO "yyyy-MM-dd"}}'
-- filter on Login Page
  AND page.path = '/customer/account/login'
```

Again: we filter on event name, date, and - in this case - on a specific page path that corresponds with the meaning of the metric.

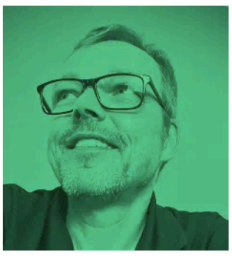
If you follow that rule, your queries will run in the most efficient way possible.

Thanks to...

The Mintminds team

Rudger de Groot

Founder & Digital Optimiser



Pascal Louwes

Co-founder & Frontend Developer

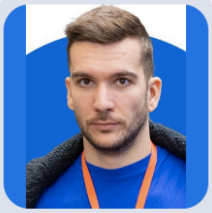
Michiel Kikkert

Cloudflare Engineer

The Superform Labs Team

Jules Stuifbergen

DATA ANALYST



Krisztián Korpa

ANALYTICS ENGINEER

Artem Korneev

ANALYTICS DEVELOPER

Simon Breton

ANALYTICS ENGINEER

Johan van de Werken

DATA EXPERT

The Social Hub

Laura Semeraro

Digital Analyst at The Social Hub. Stay curious, be bold, dare to be fun.

Links and resources

GrowthBook

- Website: https://www.growthbook.io/
- LinkedIn: https://www.linkedin.com/company/growthbook/product/

Mintminds

- Website: https://www.mintminds.com
- LinkedIn: https://www.linkedin.com/company/mintminds/
- Case Study details: https://www.mintminds.com/case/[..]dataform/

GA4Dataform by Superform Labs

- Website: https://ga4dataform.com/
- LinkedIn: https://www.linkedin.com/company/ga4dataform-superformlabs/
- Documentation: https://docs.ga4dataform.com/docs/category/get-started

The Social Hub

- Website: https://www.thesocialhub.co/
- LinkedIn: https://www.linkedin.com/company/the-social-hub-global/