

Documentation

Arooba jamil khokhar

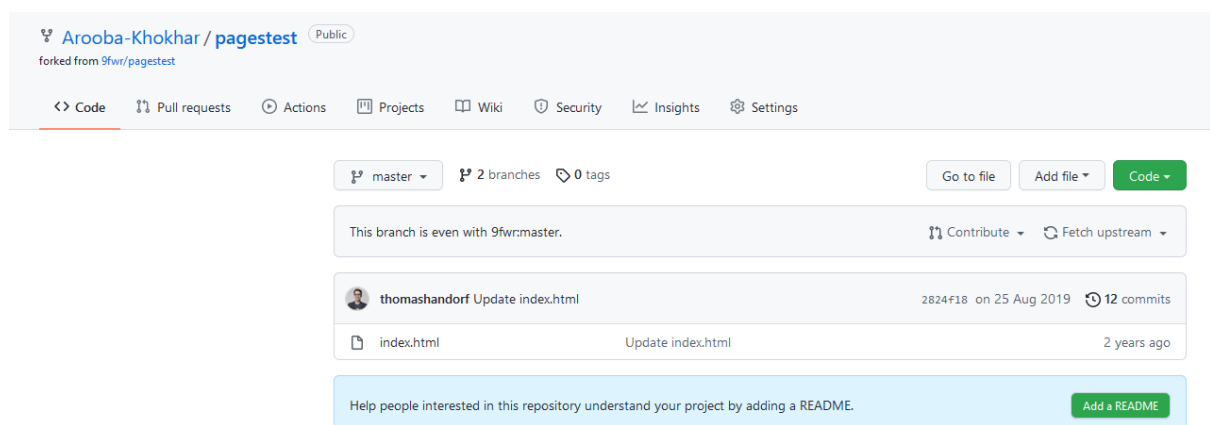
Prerequisite:

Create accounts on following sites

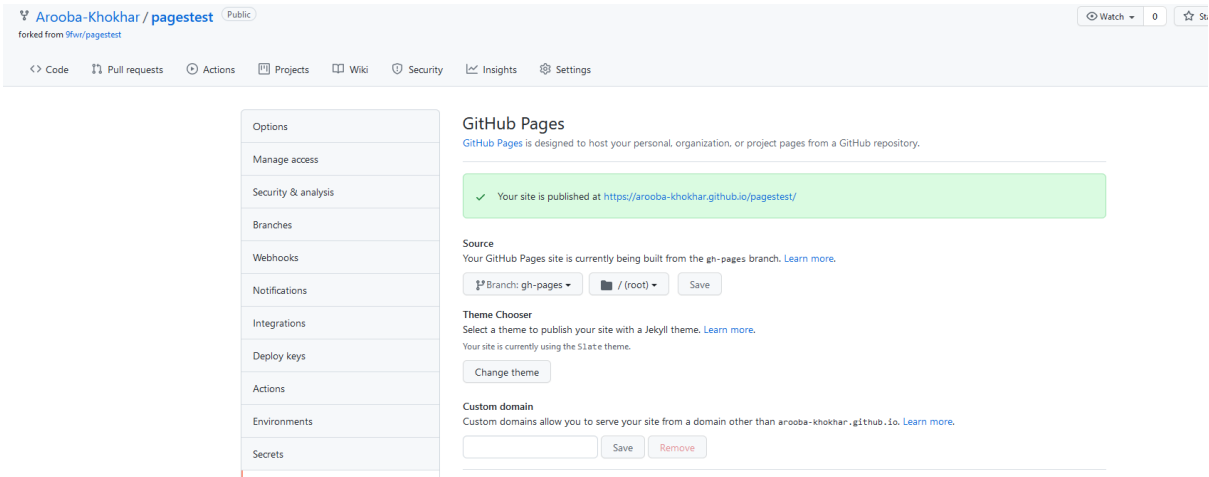
- Create account on Segment
- Create Google Analytics account
- Create Service account
- Create Bigquery account
- Configure Cloud Platform
- google analytics
- Installing Cloud SDK
- Google Data Studio

1. Data Acquisition

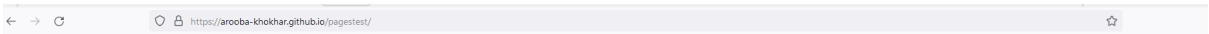
- Fork segment project



- created github page (<https://arooba-khokhar.github.io/pagestest/>)



- Github page



2. Data Warehouse

- Create account on Segment
- Add source
- Copy the Segment snippet

JS

segment Enabled
Javascript website

▼ Add Segment to your site – required

Install the Segment snippet on your website

Copy the Segment snippet and paste it high in the <head> of your website. Learn more about setting up Segment in the [Analytics.js guide](#).

```
<script>
!function(){var analytics=window.analytics=window.analytics||[];if(!analytics.initialize)if(analytics.invoked)win
analytics.load("sbp2oN04IYiRrd1DMYbTuZsMUum92NYk");
analytics.page();
})();
</script>
```

Copy Snippet

Don't have access to code? [Invite a team member!](#)

- I pasted it high in the <head> of my website.

gh-pages ▾ **pagetest / index.html** Go to file ...

Arooba-Khokhar Create index.html ✓

Latest commit 33041d0 3 days ago History

1x 1 contributor

6 lines (6 sloc) | 1.18 KB

Raw Blame Copy Edit Delete

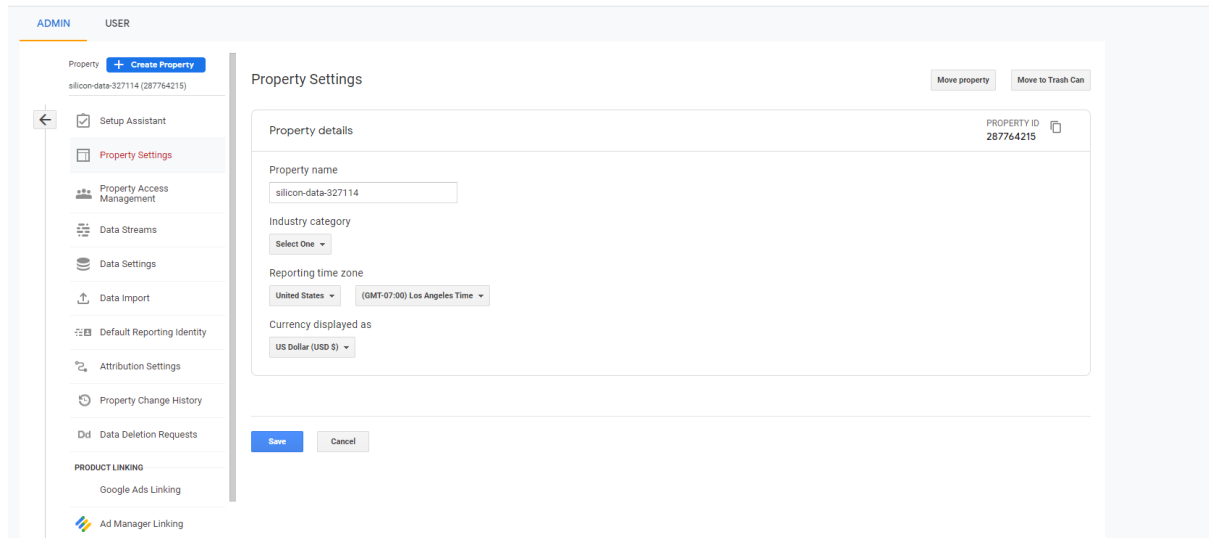
```
1 <script>
2 !function(){var analytics=window.analytics=window.analytics||[];if(!analytics.initialize)if(analytics.invoked)window.console&&console.error&&console.error("Segment
3 analytics.load("sbp2oN04IYiRrd1DMYbTuZsMUum92NYk");
4 analytics.page();
5 })();
6 </script>
```

© 2021 GitHub, Inc.

[Terms](#) [Privacy](#) [Security](#) [Status](#) [Docs](#)

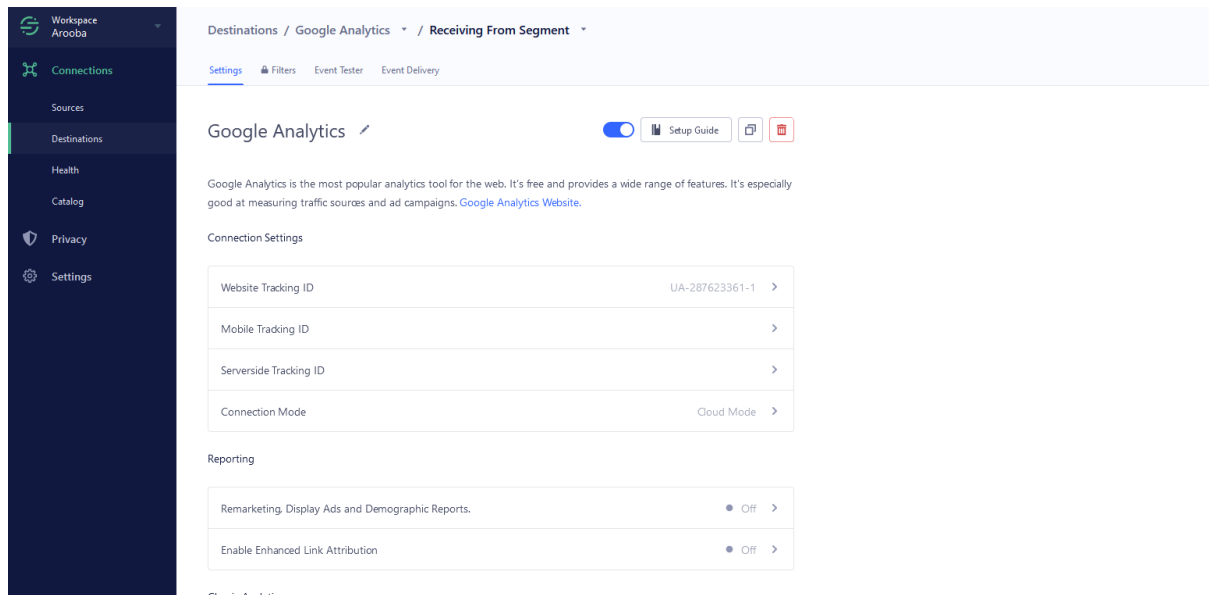
[Contact GitHub](#) [Pricing](#) [API](#) [Training](#) [Blog](#) [About](#)

- I created account on google analytics



The screenshot shows the Google Analytics 'Property Settings' page for a property named 'silicon-data-327114' (ID: 287764215). The left sidebar contains navigation links for 'ADMIN' and 'USER' views, with 'Property Settings' highlighted. The main content area is titled 'Property Settings' and includes a 'Property details' section with the following fields: 'Property name' (silicon-data-327114), 'Industry category' (Select One), 'Reporting time zone' (United States, GMT-07:00 Los Angeles Time), and 'Currency displayed as' (US Dollar (USD \$)). At the bottom of the details section are 'Save' and 'Cancel' buttons. On the far right of the settings header are 'Move property' and 'Move to Trash Can' buttons.

- Connected destination to google Analytics













The screenshot shows the Google Analytics 'Destinations' page for the 'Google Analytics' destination. The left sidebar shows the 'Connections' section with 'Destinations' selected. The main content area is titled 'Google Analytics' and includes a toggle switch (turned on) and a 'Setup Guide' button. Below this is a description of Google Analytics. The 'Connection Settings' section contains four rows: 'Website Tracking ID' (UA-287623361-1), 'Mobile Tracking ID', 'Serverside Tracking ID', and 'Connection Mode' (Cloud Mode). The 'Reporting' section contains two rows: 'Remarketing, Display Ads and Demographic Reports' and 'Enable Enhanced Link Attribution', both with a toggle switch set to 'Off'. At the bottom, there is a 'Classic Analytics' section.

- My Destination

My Destinations

Q Search destinations by name...

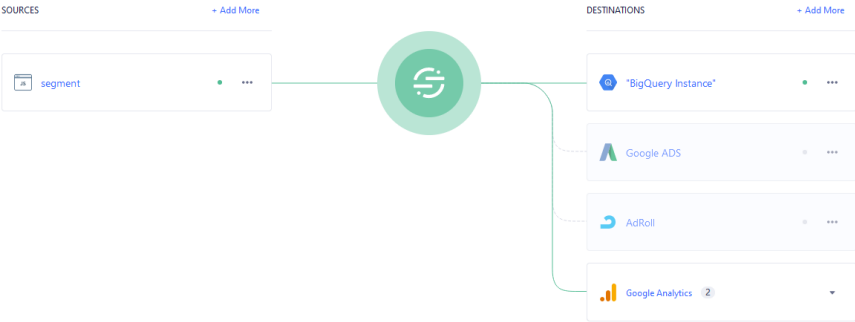
Add Destination

NAME	STATUS	CATEGORY	CREATED AT	SOURCES	
 AdRoll for segment source	Disabled	Advertising	x day ago		
 BigQuery Instance	Enabled	Warehouse	17 hours ago		
 Google for segment source	Disabled	Analytics	x day ago		
 Google ADS for segment source	Disabled	Advertising	x day ago		
 Google Analytics for segment source	Enabled	Analytics	3 hours ago		

- connected source to destination

Your Segment Overview

You've added Sources and Destinations to Segment. Add others to do even more with your data.



Sync Data ☒

Sync data between this source and warehouse. Select which collections and properties will be synced to your warehouse. Fewer collections mean faster and more frequent syncs. [See Selective Sync UI limits](#) →

COLLECTIONS	PROPERTIES
Filter by collection	Filter by property
✓ pages 16 >	✓ anonymous_id
	✓ context_ip
	✓ context_library_name
	✓ context_library_version
	✓ context_locale
	✓ context_page_path
	✓ context_page_referrer
	✓ context_page_url

- We need to track the pageview itself (every time the website loads) as well as a conversion event.

Sources / Segment

Overview Debugger Schema Settings

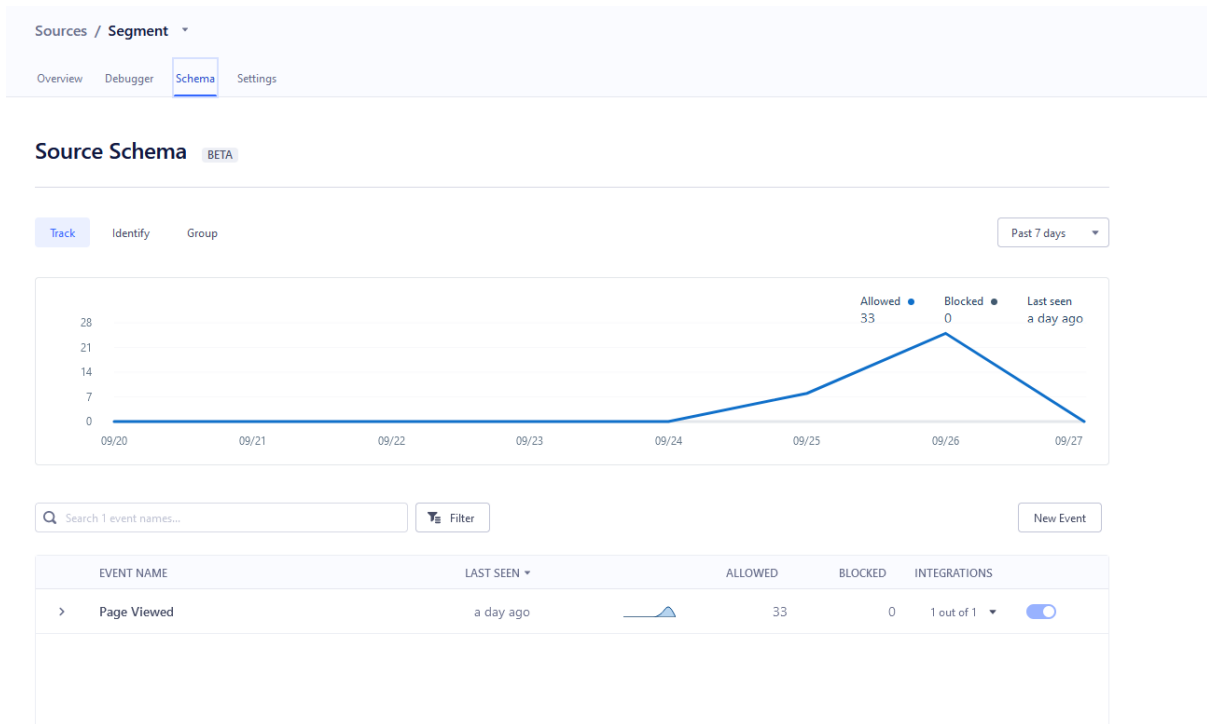
Live Pause Type to search... Advanced

✓	PAGE	/pagestest/	2021/09/26 12:49:06
✓	PAGE	/pagestest/	2021/09/26 12:49:05
✓	PAGE	/pagestest/	2021/09/26 12:49:05
✓	PAGE	/pagestest/	2021/09/26 12:49:04
✓	PAGE	/pagestest/	2021/09/26 12:49:04
✓	PAGE	/pagestest/	2021/09/26 12:49:04
✓	PAGE	/pagestest/	2021/09/26 12:48:38
✓	PAGE	/pagestest/	2021/09/26 11:41:01
✓	PAGE	/pagestest/	2021/09/26 11:41:00
✓	PAGE	/pagestest/	2021/09/26 11:41:00
✓	PAGE	/pagestest/	2021/09/26 11:40:59
✓	PAGE	/pagestest/	2021/09/26 11:40:59
✓	PAGE	/pagestest/	2021/09/26 11:40:59

Select an Event

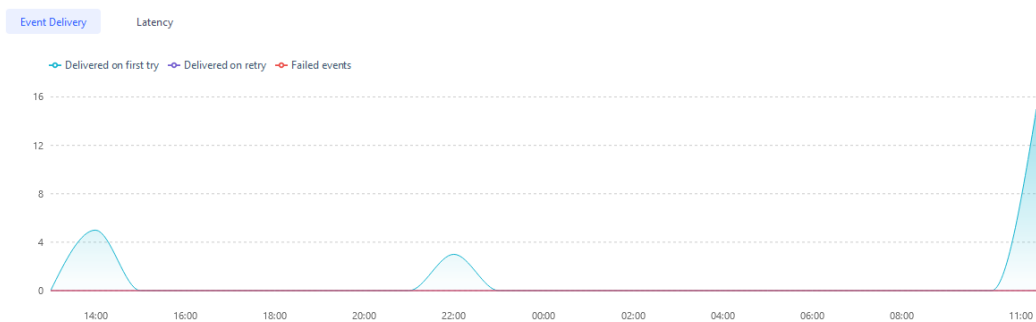
Select an Event to view the code snippet (pretty view) or complete JSON payload (raw view).

- Source schema



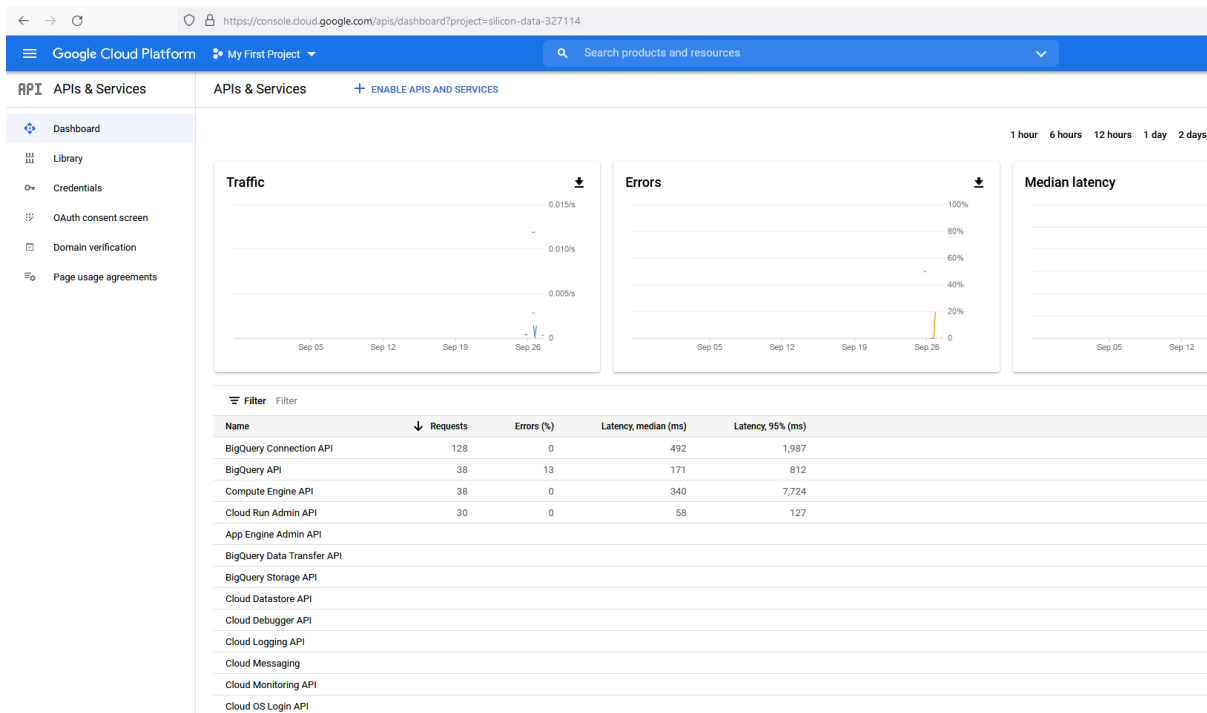
● Event Delivery


Delivery Trends



BigQuery Destination

- Create account on BigQuery
- Create a Project and Enable BigQuery





Project info

Project name
My First Project

Project ID
silicon-data-327114

Project number
655900402171

[ADD PEOPLE TO THIS PROJECT](#)

[→ Go to project settings](#)

- To re-enable billing on a project, do the following.

Google Cloud Platform

Search products and resources

Billing

MY BILLING ACCOUNTS MY PROJECTS

CREATE ACCOUNT

Filter Status: Active Enter property name or value

<input type="checkbox"/>	Billing account name	Billing account ID	Status	Last 30 days' spend	Account type	Health checks
<input type="checkbox"/>	My Billing Account	01C121-DCF8D8-858F48	Active	€0	Direct	1

- Enable the BigQuery API

Google Cloud Platform Select a project

Search products and resources

Register your application for BigQuery API in Google Cloud Platform

Google Cloud Platform allows you to manage your application and monitor API usage.

Select a project where your application will be registered

You can use one project to manage all of your applications, or you can create a different project for each application.

My First Project

Continue

- Create a Service Account for Segment
- From the Navigation panel on the left, go to IAM & admin > Service accounts

← My First Project

service account list page PERMISSIONS KEYS METRICS LOGS

Keys

⚠ Service account keys could pose a security risk if compromised. We recommend you avoid downloading service account keys and instead use the [Workload Identity Federation](#). You can learn more [here](#).

Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using [organization policies](#).
[Learn more about setting organization policies for service accounts](#)

ADD KEY ▾

Type	Status	Key	Key creation date	Key expiration date	
🔑	✔ Active	366de629f8ef9f4931e209cbf9140e65ade9738d	Sep 25, 2021	Jan 1, 10000	🗑
🔑	✔ Active	4060a9d5fb92f7fb799102d19843c96fc3b81b05	Sep 25, 2021	Jan 1, 10000	🗑

Google Cloud Platform My First Project 🔍 Search products and resources ▾

IAM & Admin

← civil-willow-327114

DETAILS PERMISSIONS KEYS METRICS LOGS

Keys

⚠ Service account keys could pose a security risk if compromised. We recommend you avoid downloading service account keys and instead use the [Workload Identity Federation](#). You can learn more about the best way [here](#).

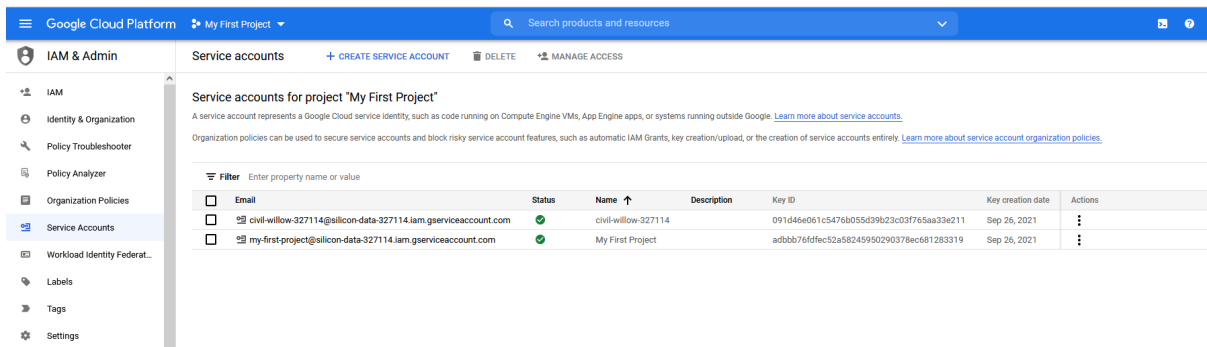
Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using [organization policies](#).
[Learn more about setting organization policies for service accounts](#)

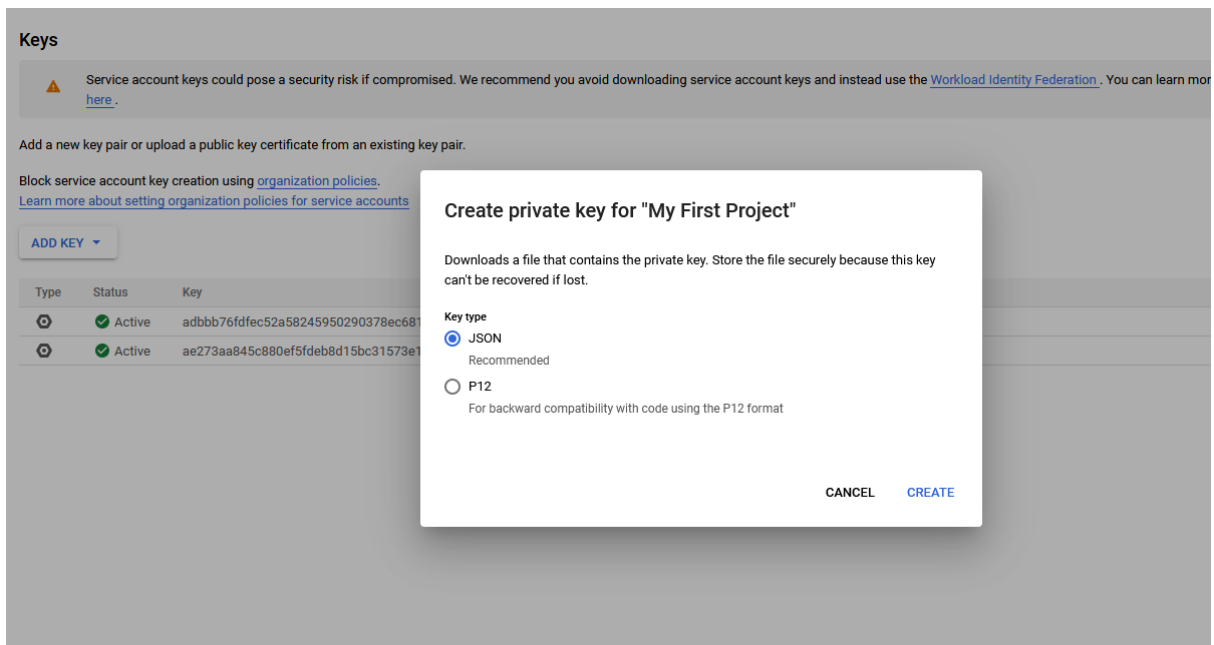
ADD KEY ▾

Type	Status	Key	Key creation date	Key expiration date	
🔑	✔ Active	091d46e061c5476b055d39b23c03f765aa33e211	Sep 26, 2021	Jan 1, 10000	🗑

- Create Service Account



- Create a JSON key



- Create the Warehouse in Segment

Connection

Project ID *

silicon-data-327114

Location

Credentials *

```

/v1/certs",
  "client_x509_cert_url": "https://www.googleapis.com/robot/v1/metadata/x509/my-first-project%40silicon-data-327114.iam.gserviceaccount.com"
}

```

 Type to search...

SCHEMA	SYNC SCHEDULE ▾	STATUS
segment Updated, Sep 26 2021, 5:50 PM	Last, yesterday at 5:50 PM Next, today at 5:53 PM	● Complete 33 rows synced

- Warehouse Overview

We're syncing your warehouse for the first time. Depending on the amount of data you've sent to Segment, the initial sync can take a few minutes up to 24 hours.

Warehouse Overview

🔍 Type to search...

✔ We're syncing your warehouse for the first time. Depending on the amount of data you've sent to Segment, the initial sync can take a few minutes up to 24 hours.

SCHEMA	SYNC SCHEDULE ▾	STATUS
segment		First sync pending

Sources / Segment

OverviewDebuggerSchemaSettings

segment

JavaScript website

Enabled

Add Segment to your site – required

Destinations

2 Enabled, 3 Disabled

Add Destination

NAME	STATUS
AdRoll Advertising	Disabled
BigQuery Instance BigQuery	Enabled
Google Analytics	Disabled
Google ADS Advertising	Disabled
Google Analytics Analytics	Enabled

Sources / Segment

OverviewDebuggerSchemaSettings

LivePause

Type to search...

Advanced

	PAGE	/pagetest/	2021/09/25 14:52:09
	PAGE	/pagetest/	2021/09/25 14:52:08
	PAGE	/pagetest/	2021/09/25 14:52:07
	PAGE	/pagetest/	2021/09/25 14:52:06
	PAGE	/pagetest/	2021/09/25 14:48:15

Select an Event

Select an Event to view the code snippet (pretty view) or complete JSON payload (raw view).

My Sources

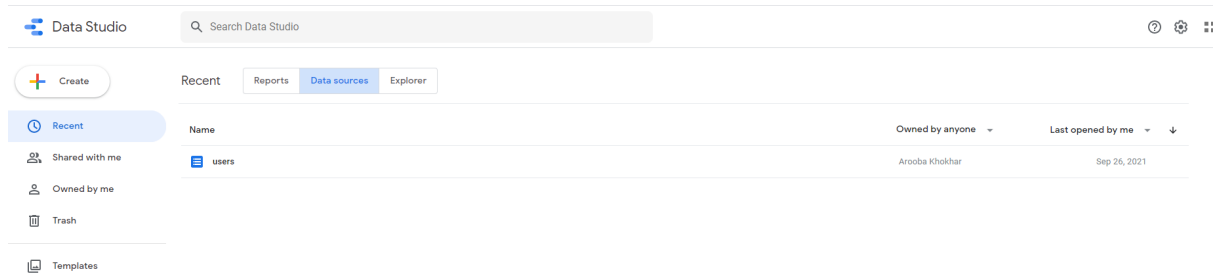
Search sources by name...

Add Source

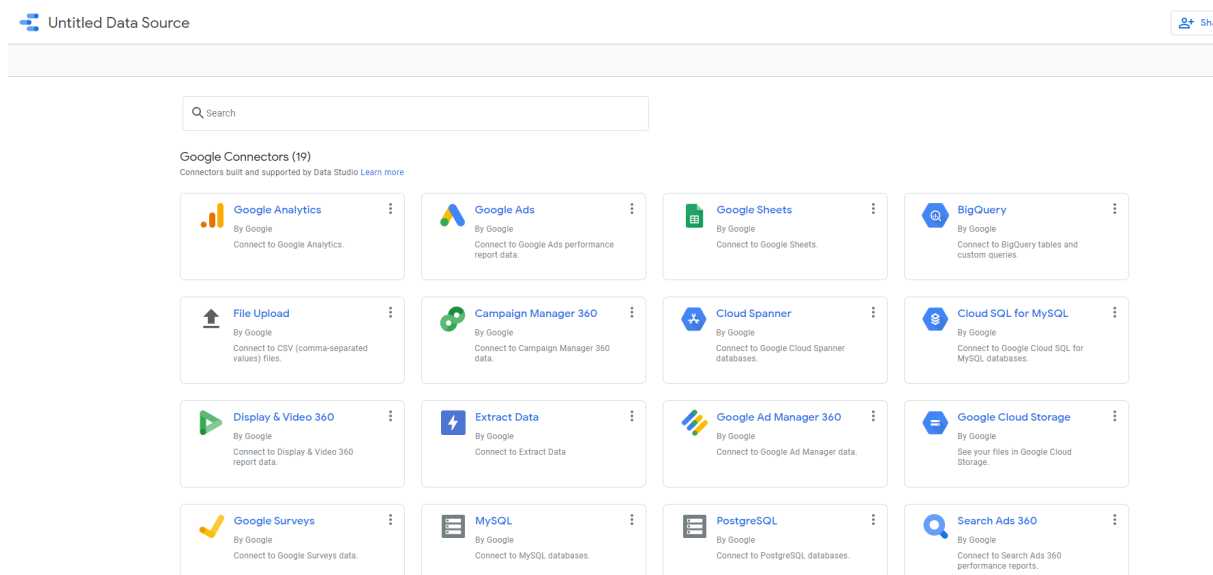
NAME	STATUS	CATEGORY	DESTINATIONS	ENVIRONMENT
segment	Enabled	Website	+2	-

3. Visualize


- Create an account with Google Data Studio




- Select “Bigquery”



- the custom query, I chose my GCP project and then type my query

 Make your BigQuery reports load even faster with BigQuery BI Engine. [Learn More](#)

**BigQuery**
By Google
BigQuery is Google's fully managed, petabyte scale, low-cost analytics data warehouse. BigQuery charges for querying/processing of data. Those queries are charged to the credit card of the billing project.
[LEARN MORE](#) [REPORT AN ISSUE](#)

RECENT PROJECTS

MY PROJECTS

SHARED PROJECTS

CUSTOM QUERY

PUBLIC DATASETS

Billing Project

My First Project


My First Project

My Project

My Project

Enter Custom Query

```
1 SELECT DATE_TRUNC('day', derived_timestamp) as "Date",
2 COUNT(DISTINCT(domain_userid)) as "Uniques"
3 COUNT(*) AS "page_views"
4 FROM segment.accounts
5 WHERE derived_timestamp > current_date - integer '31'
6 GROUP BY 1
7 ORDER BY 1;
8 |
```

☒ Use Legacy SQL 

Parameters are not supported in legacy SQL mode

Create a dashboard with

- a graph showing the number of unique users per day and pageviews per day
- show the user based daily conversion rate over time. i.e. the percentage of users that clicked to button for which you created the event on each day

Add data to report

Data credentials:

Make your BigQuery reports load even faster with BigQuery BI Engine.
[Learn More](#)

BigQuery

By Google

BigQuery is Google's fully managed, petabyte scale, low-cost analytics data warehouse. BigQuery charges for querying/processing of data. Those queries are charged to the credit card of the billing project.

[LEARN MORE](#)
[REPORT AN ISSUE](#)

RECENT PROJECTS

MY PROJECTS

SHARED PROJECTS

CUSTOM QUERY

PUBLIC DATASETS

Billing Project

My First Project

My First Project

Enter Custom Query

```

1 SELECT DATE_TRUNC('day', derived_tstamp) as "Date",
2 COUNT(DISTINCT(domain_userid)) as "Uniques"
3 COUNT(*) AS "page_views"
4 FROM "atomic"-events
5 WHERE derived_tstamp > current_date - integer '31'
6 GROUP BY 1
7 ORDER BY 1;

```

☒ Use Legacy SQL

Parameters are not supported in legacy SQL mode

Part B

SQL :

- Created Customer table

```
CREATE DATABASE Customers;
```

```
CREATE TABLE Customers
```

```
( Emp_ID INT,
```

```
  Name VARCHAR(20),
```

```
  City VARCHAR(20) ,
```

```
  Age INT,
```

```
  Profession VARCHAR(20),
```

```
);
```

- Add fields (NAME,CITY,PROFESSION,AGE)

```
INSERT INTO Customers
```

```
(Name, City, Age,Profession )
```

```
VALUES
```

```
('Russell', 'berlin', '12', 'doctor'),  
( 'Richard', 'munich', '13', 'lawyer'),  
( 'Rusty', 'hamburg', '15', 'actor'),  
( 'Lexi', 'halle', '19', 'director');
```

- Add new column “LOAD DATE” with the current timestamp

```
ALTER TABLE Customers  
ADD LoadDate datetime;
```

- Return list of cities and the number of customers

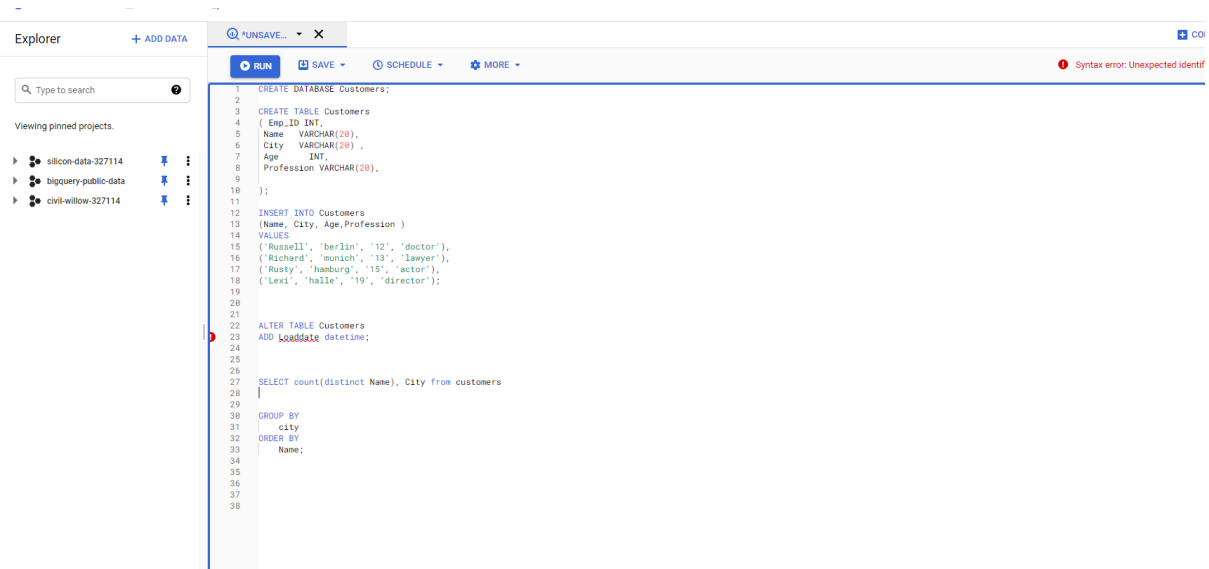
```
SELECT count(distinct Name), City from customers
```

```
GROUP BY
```

```
    city
```

```
ORDER BY
```

```
    Name;
```



ETL/python

I unpivoted form where rows are facts with the dimensions, country and date.

```

12 pd.drop(['Province/State', 'Lat', 'Long'], axis=1, inplace=True)
13 pd = pd.rename(columns={"Country/Region": "Country"})
14 # do the unpivoting here ...
15
16
17 pd = df.melt(id_vars=['Country'], var_name='Date')
18 pd
19
20
21 # ignore BigQuery Loading for this test
22 # client = bigquery.Client()
23 # table_id='f-26914.test_covid.confirmed'
24 # job_config = bigquery.LoadJobConfig(
25 #     write_disposition="WRITE_TRUNCATE"
26 # )
27 # job = client.load_table_from_dataframe(
28 #     pd,
29 #     table_id,
30 #     job_config=job_config,
31 #     location="europe-west3",
32 # )
33 # job.result()
34 main()

```

```

Console 1/A
In [19]: pd = df.melt(id_vars=['Country'], var_name='Date')
Traceback (most recent call last):

  File "<ipython-input-19-553b53d75c0a>", line 1, in <module>
    pd = df.melt(id_vars=['Country'], var_name='Date')
NameError: name 'df' is not defined

In [20]: pd = pd.melt(id_vars=['Country'], var_name='Date')

In [21]: pd
Out[21]:
   Country      Date  value
0  Afghanistan  1/22/20     0
1    Albania    1/22/20     0
2    Algeria    1/22/20     0
3   Andorra    1/22/20     0
4    Angola    1/22/20     0
...
171022  Vietnam  9/25/21  746678
171023  West Bank and Gaza  9/25/21  395677
171024    Yemen    9/25/21   8891
171025    Zambia    9/25/21  208829
171026    Zimbabwe    9/25/21  129505

[171027 rows x 3 columns]

In [22]:

```