

Serverless iot data processing

Steps:

- Start the data pipeline
- Check that data is flowing
- Create a datastudio dashboard
- Final output

Start the data pipeline

Might need to enable Compute API

Data streaming from a Raspberry Pi

If you constructed a Raspberry Pi IoT weather sensor, start the script that will read the weather data and push it to Google Cloud Pub/Sub. If you aren't in the `/home/pi/iot-data-pipeline` directory, move there first

```
cd /home/pi/iot-data-pipeline
```

Start the weather script

```
python checkWeather.py
```

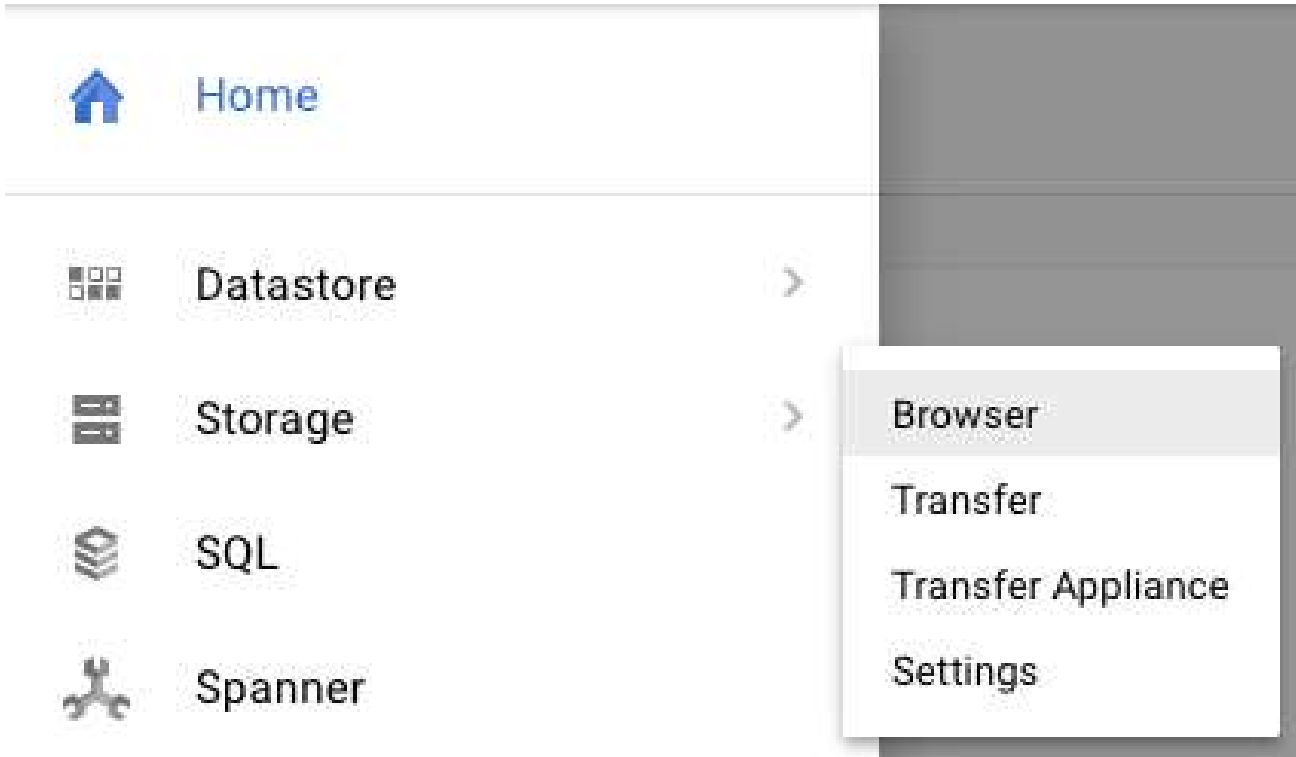
You should see the terminal window echo the weather data results every minute. With data flowing, you can skip to the next section (Check that Data is Flowing).

Simulated data streaming

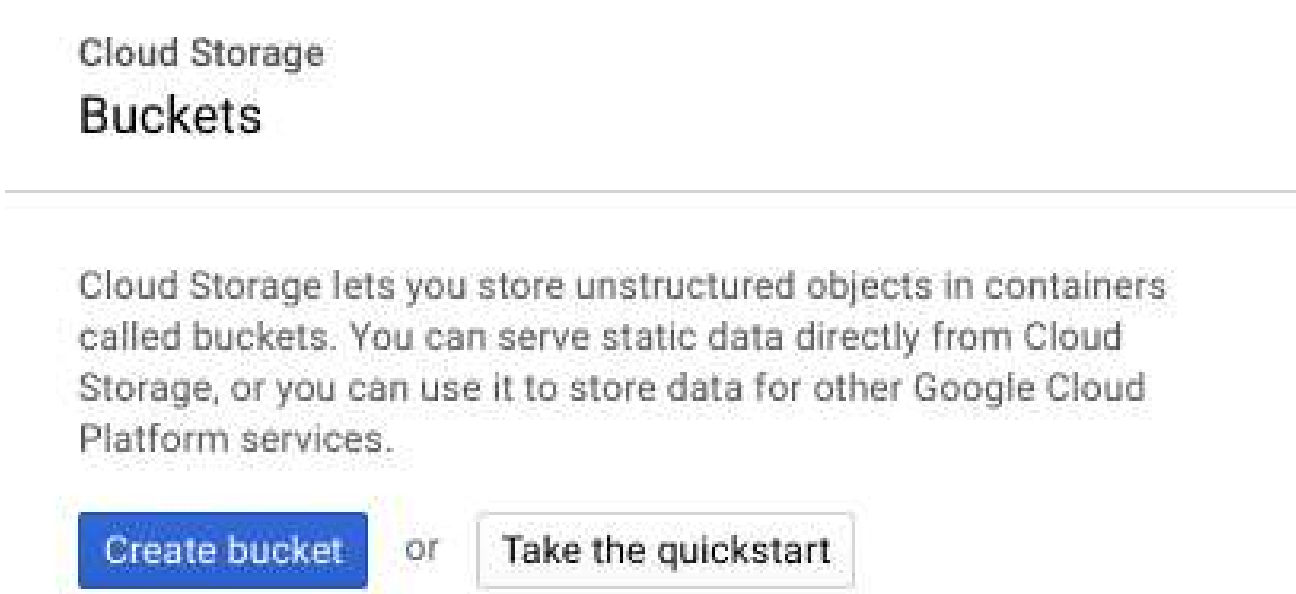
If you didn't build the IoT weather sensor, you can simulate data streaming by using a public dataset that has been stored in Google Cloud Storage and feeding it into the existing Pub/Sub topic. Google Dataflow along with a Google-provided template for reading from Cloud Storage and publishing to Pub/Sub will be used.

As part of the process, Dataflow will need a temporary storage location, so let's create a storage bucket for this purpose.


From the Cloud Console, select Storage and then Browser.





Click the Create Bucket button





Choose a name for the storage bucket (remember, it must be a name that is globally unique across all of Google Cloud) and click on the Create button. Remember the name of this storage bucket as it will be needed shortly.

Storage


Browser

Transfer

Transfer Appliance


Settings

← Create a bucket

Name 

Must be unique across Cloud Storage. **Privacy:** Do not include sensitive information in your bucket name. Others can discover your bucket name if it matches a name they're trying to use.

iot2analytics-tmp

Default storage class 

☒ **Multi-Regional**
Use to stream videos and host hot web content.
Best for data accessed frequently around the world.


☐ **Regional**
Use to store data and run data analytics.
Best for data accessed frequently in one part of the world.

☐ **Nearline**
Use to store rarely accessed documents.
Best for data accessed less than once per month.

☐ **Coldline**
Use to store very rarely accessed documents.
Best for data accessed less than once per year.

Multi-Regional location
Redundant across 2+ regions within your selected location.

United States

 Specify labels

Create

Cancel

From the Cloud Console, select Dataflow.



Google Cloud Platform



Home



Pub/Sub



Dataproc



Dataflow

Click on Create Job from Template (upper portion of the screen)



Dataflow

Jobs

[+ CREATE JOB FROM TEMPLATE](#)

Big Data Cloud Dataflow

Cloud Dataflow provides scalable data-processing pipelines for small and large jobs. Use the [Cloud Dataflow SDK](#) to define jobs, and then monitor them here.

[Try Dataflow](#)

or

[Learn more](#)

Fill in the job details as shown below paying attention to the following:

- Enter a job name of dataflow-gcs-to-pubsub
- Your region should auto-select according to where your project is hosted and should not need to be changed.
- Select a Cloud Dataflow template of GCS Text to Cloud Pub/Sub
- For the Input Cloud Storage File(s), enter gs://codelab-iot-data-pipeline-sampleweatherdata/*.json (this is a public dataset)
- For the Output Pub/Sub Topic, the exact path will depend upon your project name and will look something like "projects/yourProjectName/topics/weatherdata"
- Set the Temporary Location to the name of the Google Cloud Storage bucket you just created **along with a filename prefix of "tmp"**. It should look like "gs://myStorageBucketName/tmp".

When you have all the information filled in (see below), click the Run job button

**Job name**Specify a job name for the Dataflow job. [?](#)

dataflow-gcs-to-pubsub

Cloud Dataflow Regional EndpointThe regional endpoint where metadata will be stored for the job. [Learn more](#)

us-central1

Cloud Dataflow template

A pipeline that reads a text file stored in GCS and outputs each line to a Pub/Sub topic.

GCS Text to Cloud PubSub

Parameters**Input Cloud Storage File(s)**

Path of the file pattern glob to read from. ex: gs://bucket-name/path/*.csv

gs://code-lab-iot-data-pipeline-sample-weatherdata/*.json

Output Pub/Sub Topic

The name of the topic which data should be published to. The name should be in the format of projects/<project-id>/topics/<topic-name>.

projects/iot2analytics/topics/weatherdata

Temporary Location

Path and filename prefix for writing temporary files. ex: gs://MyBucket/tmp

gs://iot2analytics-tmp/tmp/

Runtime Environment**Max workers** (Optional)

The maximum number of Google Compute Engine instances to be made available to your pipeline during execution, must be larger than 0.

Zone (Optional)The Compute Engine availability zone for launching worker instances to run your pipeline. (<https://cloud.google.com/compute/docs/regions-zones/regions-zones>)**Service account email** (Optional)

The email address of the service account to run the job as.

Machine type (Optional)

The machine type for Google Compute Engine instances used in your pipeline execution. E.g., n1-standard-1.

Additional parameters

Run job

Cancel

The Dataflow job should start to run.

dataflow-gcs-to-pubsub

LOGS

Read Text Data

Running

Write to PubSub

Running

Job

Job summary

Job name

dataflow-gcs-to-pubsub

Job ID

2018-02-01_08_54_59-14441071159444143835

Region

us-central1

Job status

Running

Stop job

SDK version

Apache Beam SDK for Java 2.2.0

Job type

Batch

Start time

Feb 1, 2018, 9:55:00 AM

Elapsed time

40 sec

Autoscaling

...

Resource metrics

Current vCPUs

0

Total vCPU time

0 vCPU hr

Current memory

0 B

Total memory time

0 GB hr

Current PD

0 B

Total PD time

0 GB hr

Current SSD PD

0 B

Total SSD PD time

0 GB hr

Total shuffle usage

-

It should take approximately a minute for the Dataflow job to complete.

```
at emitTwo (events.js:186:13)
at Request.emit (events.js:191:7)
at Request.<anonymous> (/user_code/node_modules/@google-cloud/bigquery/node_modules/request/request.js:1163:10)
at emitOne (events.js:196:13)
at Request.emit (events.js:188:7)
code: 400,
errors:
  [ { message: 'Invalid project ID \'myProject\''. Project IDs must contain 6-63 lowercase letters, digits, or dashes. IDs m
ust start with a letter and may not end with a dash.',
    domain: 'global',
    reason: 'invalid' } ],
  response: undefined,
  message: 'Invalid project ID \'myProject\''. Project IDs must contain 6-63 lowercase letters, digits, or dashes. IDs must s
tart with a letter and may not end with a dash.' }
andred@iot2analytics-183920:~$
```

Check that data is flowing

Cloud Function logs

Ensure that the Cloud Function is being triggered by Pub/Sub

gcloud beta functions logs read function-weatherPubSubToBQ

The logs should show that the function is executing, data is being received and that it is being inserted into BigQuery

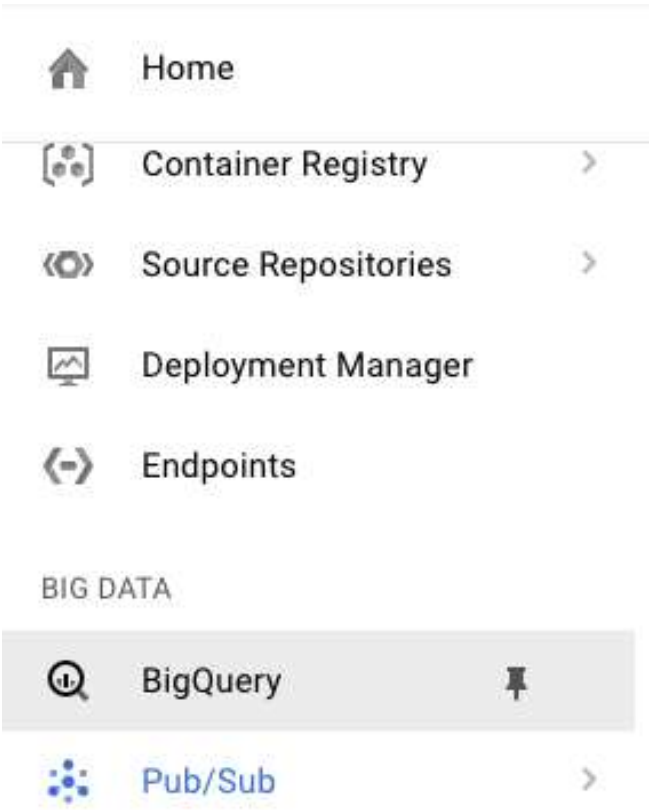
```

function-weatherPubSubToBQ 28796420510154 2018-01-30 20:01:18.750 Function execution started
function-weatherPubSubToBQ 28796420510154 2018-01-30 20:01:19.878 Incoming data: [object Object]
function-weatherPubSubToBQ 28796420510154 2018-01-30 20:01:19.923 Function execution took 1174 ms, finished with status: 'ok'
function-weatherPubSubToBQ 28796420510154 2018-01-30 20:01:24.351 Inserted:
function-weatherPubSubToBQ 28796420510154 2018-01-30 20:01:25.051 { pressure: '24.19',
                                                                    temperature: '70.90',
                                                                    dewpoint: '41.59',
                                                                    timecollected: '2018-01-30 20:01:11',
                                                                    latitude: '37.421655',
                                                                    sensorID: 's-testing',
                                                                    zipcode: '94043',
                                                                    longitude: '-122.085637',
                                                                    humidity: '18.58' }

```

BigQuery data

Check to make sure that data is flowing into the BigQuery table. From the Cloud Console, go to BigQuery (bigquery.cloud.google.com).



Under the project name (on the left hand side of the window), click on the Dataset (weatherData), then on the table (weatherDataTable) and then click on the Query Table button

COMPOSE QUERY

Query History

Job History

Filter by ID or label



lot2Analytics



▼ weatherData

■ weatherDataTable

▼ Public Datasets

▶ bigquery-public-data:hacker_news

▶ bigquery-public-data:noaa_gsod

▶ bigquery-public-data:samples

▶ bigquery-public-data:usa_names

▶ gdelt-bq:hathitrustbooks

▶ gdelt-bq:internetarchivebooks

▶ lookerdata:cdc

▶ nyc-tlc:green

Table Details: weatherDataTable

Refresh

Query Table

Copy Table

Export Table

Schema Details Preview

sensorID	STRING	NULLABLE	Describe this field...
timecollected	TIMESTAMP	NULLABLE	Describe this field...
zipcode	INTEGER	NULLABLE	Describe this field...
latitude	FLOAT	NULLABLE	Describe this field...
longitude	FLOAT	NULLABLE	Describe this field...
temperature	FLOAT	NULLABLE	Describe this field...
humidity	FLOAT	NULLABLE	Describe this field...
dewpoint	FLOAT	NULLABLE	Describe this field...
pressure	FLOAT	NULLABLE	Describe this field...

Add New Fields

Add an asterisk to the SQL statement so it reads `SELECT * FROM...` as shown below and then click the RUN QUERY button

New Query ?

Query Editor

UDF Editor



```
1 SELECT * FROM [iot2analytics:weatherData.weatherDataTable] LIMIT 1000
```

SQL

Ctrl + Enter: run query, Tab or Ctrl + Space: autocomplete.

RUN QUERY ▼

Save Query

Save View

Format Query

Show Options



If prompted, click on the Run query button

Confirm query

With this query, you will be billed for all the data in the table (even if your query contains a LIMIT clause). If you're using the free tier, this query still counts against your free quota.

You can use table preview instead to see records for free and without affecting quotas.

☐ Don't show this again

Run query

Go to table preview

Cancel

If you see results, then data is flowing properly.

New Query ?

Query Editor UDF Editor



```
1 SELECT * FROM [iot2analytics:weatherData.weatherDataTable] LIMIT 1000
```

SQL

Ctrl + Enter: run query, Tab or Ctrl + Space: autocomplete.

RUN QUERY

Save Query

Save View

Format Query

Show Options

Query complete (1.2s elapsed, 0 B processed)



Results

Details

Download as CSV

Download as JSON

Save as Table

Save to Google Sheets

Row	sensorID	timecollected	zipcode	latitude	longitude	temperature	humidity	dewpoint	pressure
1	s-testing	2018-01-30 20:02:11.000 UTC	94043	37.421655	-122.085637	70.86	18.33	41.46	24.19
2	s-testing	2018-01-30 20:11:13.000 UTC	94043	37.421655	-122.085637	70.85	18.24	41.42	24.18
3	s-testing	2018-01-30 20:06:12.000 UTC	94043	37.421655	-122.085637	70.89	18.28	41.47	24.19
4	s-testing	2018-01-30 20:09:13.000 UTC	94043	37.421655	-122.085637	70.86	18.23	41.43	24.18
5	s-testing	2018-01-30 20:08:12.000 UTC	94043	37.421655	-122.085637	70.88	18.27	41.46	24.18
6	s-testing	2018-01-30 20:15:14.000 UTC	94043	37.421655	-122.085637	70.87	18.26	41.45	24.18
7	s-testing	2018-01-30 20:01:11.000 UTC	94043	37.421655	-122.085637	70.9	18.58	41.59	24.19
8	s-testing	2018-01-30 20:14:14.000 UTC	94043	37.421655	-122.085637	70.83	18.22	41.39	24.18
9	s-testing	2018-01-30 20:13:14.000 UTC	94043	37.421655	-122.085637	70.86	18.25	41.43	24.18

Table JSON

First < Prev Rows 1 - 9 of 16 Next > Last

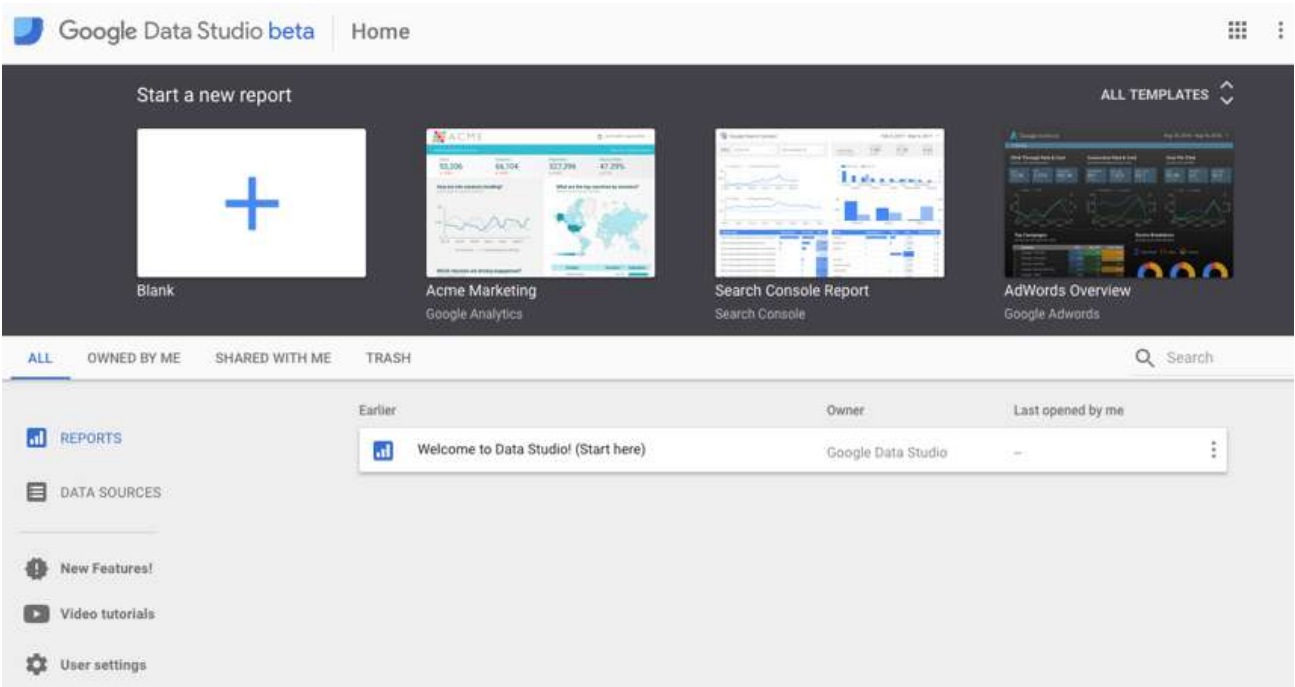
With data flowing, you are now ready to build out an analytics dashboard.

Create a Data Studio dashboard

Google Data Studio turns your data into informative dashboards and reports that are easy to read, easy to

share, and fully customizable.

From your web browser, go to <https://datastudio.google.com>



Under "Start a new report", click on Blank and then click on the Get Started button



Welcome



Terms



Preferences



Welcome to Google Data Studio

Turn your data into informative dashboards and reports that are easy to read, easy to share, and fully customizable. Data Studio allows you to tell great data stories to support better business decisions.

GET STARTED



Connect

Easily connect to all your data source. Connect to multiple Data Sources within one report.



Visualize

Select a variety of visualizations. Custom layout and themes. Apply dimensions and metrics. Create custom metrics.



Share

Easy sharing. Individuals, groups of users, public. Realtime collaboration.

Click the checkbox to accept the terms, click the Next button, select which emails you are interested in receiving and click on the Done button. Once again, under "Start a new report", click on Blank

1

2

3

X

WelcomeTermsPreferences

Please agree to the terms and conditions

To use Google Data Studio you must first accept the terms of service.

English

Google Data Studio Terms of Service

By using Google Data Studio (the "Service"), you agree to the following terms in addition to the Google Terms of Service ("Google ToS") available at www.google.com/policies/terms/ (or at such other URL as Google may provide).

1. Services.

1.1 Facilities and Data Transfer. All facilities used to store and process Customer Data will adhere to reasonable security standards no less protective than the security standards at facilities where Google stores and processes its own information of a similar type. Google has implemented at least industry standard systems and procedures to ensure the security and confidentiality of Customer Data, protect against anticipated threats or hazards to the security or integrity of Customer Data and protect against unauthorized access to or use of Customer Data. As part of providing the Service, Google may transfer, store and process Customer Data in the United States or any other country in which Google or its agents maintain facilities. Because the Service was created in this transfer, processing and storage of Customer

☐ I acknowledge I have read and agree to the above Google Data Studio Additional Terms.

PREVIOUS

NEXT

Click on the Create New Data Source button










Add a data source

A data source provides data for charts. Select an existing data source or click CREATE NEW DATA SOURCE.

[OKAY, GOT IT](#)

Select Data Source



-  [Sample] World Population Data 2...
-  [Sample] Google Analytics Data
-  [Sample] Firebase Analytics Data (...)
-  [Sample] Firebase Analytics Data ...
-  [Sample] Firebase Analytics Data (...)
-  [Sample] AdWords Data
-  [Sample] YouTube Data
-  [Sample] Rio Olympics Data
-  [Sample] Search Console Data (Sit...

[CREATE NEW DATA SOURCE](#)

Click on BigQuery, then on the Authorize button and then choose the Google account you wish to use with Data Studio (it should be the same one that you have been using for the codelab).

Untitled Data Source

Field Editing in Reports: ONCANCELCONNECT

Connectors

- File Upload
- AdWords
- Attribution 360
- BigQuery
- Cloud SQL
- DCM
- DFP
- Google Cloud Storage
- Google Analytics
- Google Sheets
- MySQL
- PostgreSQL
- Search Console
- YouTube Analytics

BigQuery

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

Authorization

Data Studio requires authorization to connect to your BigQuery projects.

AUTHORIZE

Click on the Allow button

Google Data Studio wants to

View your data in Google BigQuery

Allow Google Data Studio to do this?

By clicking Allow, you allow this app to use your information in accordance to their terms of service and privacy policies. You can remove this or any other app connected to your account in [My Account](#)

CANCEL

ALLOW

Select your project name, dataset and table. Then click the Connect button.

Untitled Data Source

Field Editing in Reports: ONCANCELCONNECT

Connectors

- File Upload
- AdWords
- Attribution 360
- BigQuery
- Cloud SQL

BigQuery

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

	Project	Dataset	Table
MY PROJECTS			
SHARED PROJECTS	lot2Analytics	weatherData	weatherDataTable
	My First Project		
CUSTOM QUERY			
PUBLIC DATASETS			

Change the type fields as shown below (everything should be a number except for timecollected and sensorID). Note that timecollected is set to Date Hour (and not just Date). Change the Aggregation fields as shown below (dewpoint, temperature, humidity and pressure should be averages and everything else should be set to "None"). Click on the Create Report button.

weatherDataTable

Field Editing in Reports: ON USING OWNER'S CREDENTIALS CREATE REPORT

EDIT CONNECTION

Index	Field	Type	Aggregation	Description
1	zipcode	123 Number	None	
2	dewpoint	123 Number	Average	
3	timecollected	Date Hour (YYYYMMDD...	None	
4	latitude	123 Number	None	
5	temperature	123 Number	None	
6	humidity	123 Number	Average	
7	pressure	123 Number	Average	
8	sensorID	RBC Text	None	
9	longitude	123 Number	Average	

Confirm by clicking the Add to report button

You are about to add a data source to this report



Note that **Report Editors** can create charts using the new data source(s), and can add dimensions and metrics not currently included in the report.

CANCEL

ADD TO REPORT

If asked to select your Google account, do so and then click the Allow button to let Data Studio store its reports in Google Drive.

Google Data Studio wants to



View and manage the files in your Google Drive



Allow Google Data Studio to do this?

By clicking Allow, you allow this app to use your information in accordance to their terms of service and privacy policies. You can remove this or any other app connected to your account in [My Account](#)

CANCEL

ALLOW

You are presented with a blank canvas on which to create your dashboard. From the top row of icons, choose Time Series.



Untitled Report

File Edit View Insert Page Arrange

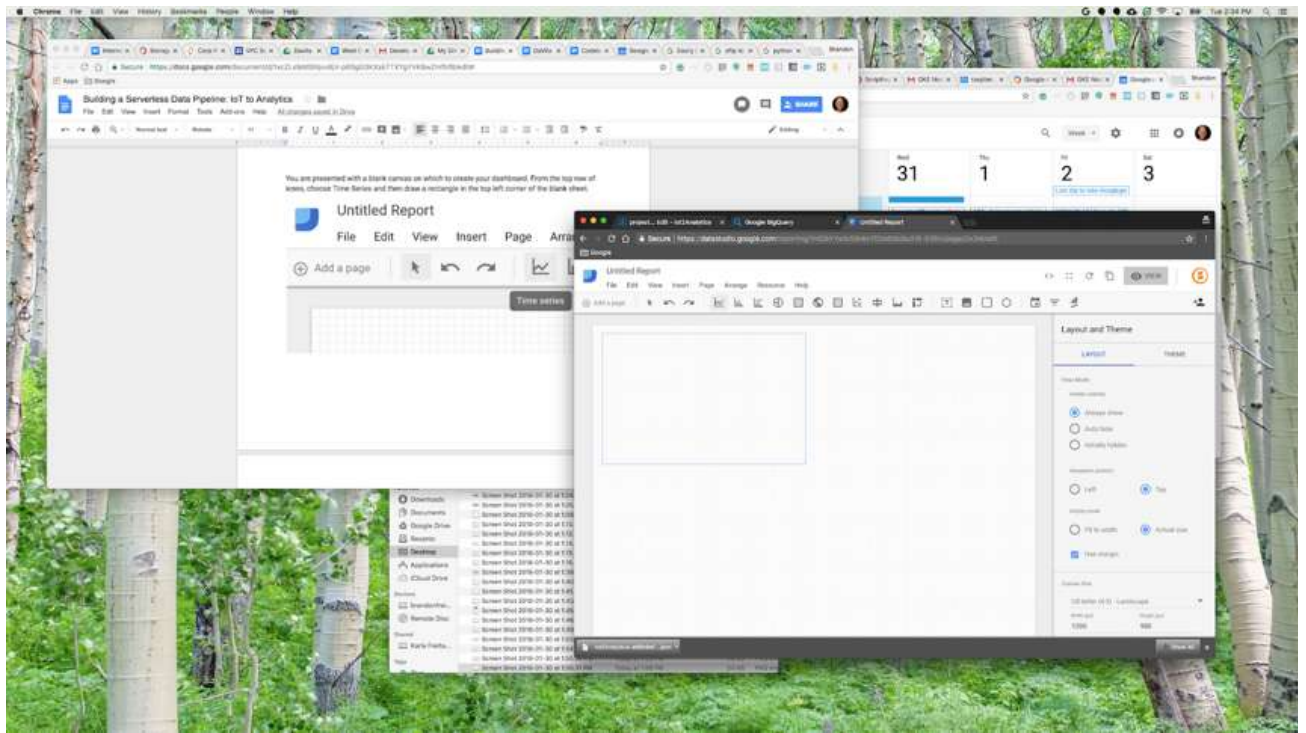


Add a page



Time series

Draw a rectangle in the top left corner of the blank sheet. It should occupy about ¼ of the total blank sheet.



On the right hand side of the window, select the Style tab. Change Missing Data from "Line To Zero" to "Line Breaks". In the Left Y-Axis section, delete the 0 from Axis Min to change it to (Auto).

Time series Properties

DATA

STYLE

Series #1

☒ Line

☐ Bars

2

☐ Cumulative

☐ Show Points

☐ Show data labels

Trendline

None

Missing Data

Line To Zero

Axes

☒ Show axes

Left Y-Axis

☐ Show axis title

Axis Min

0

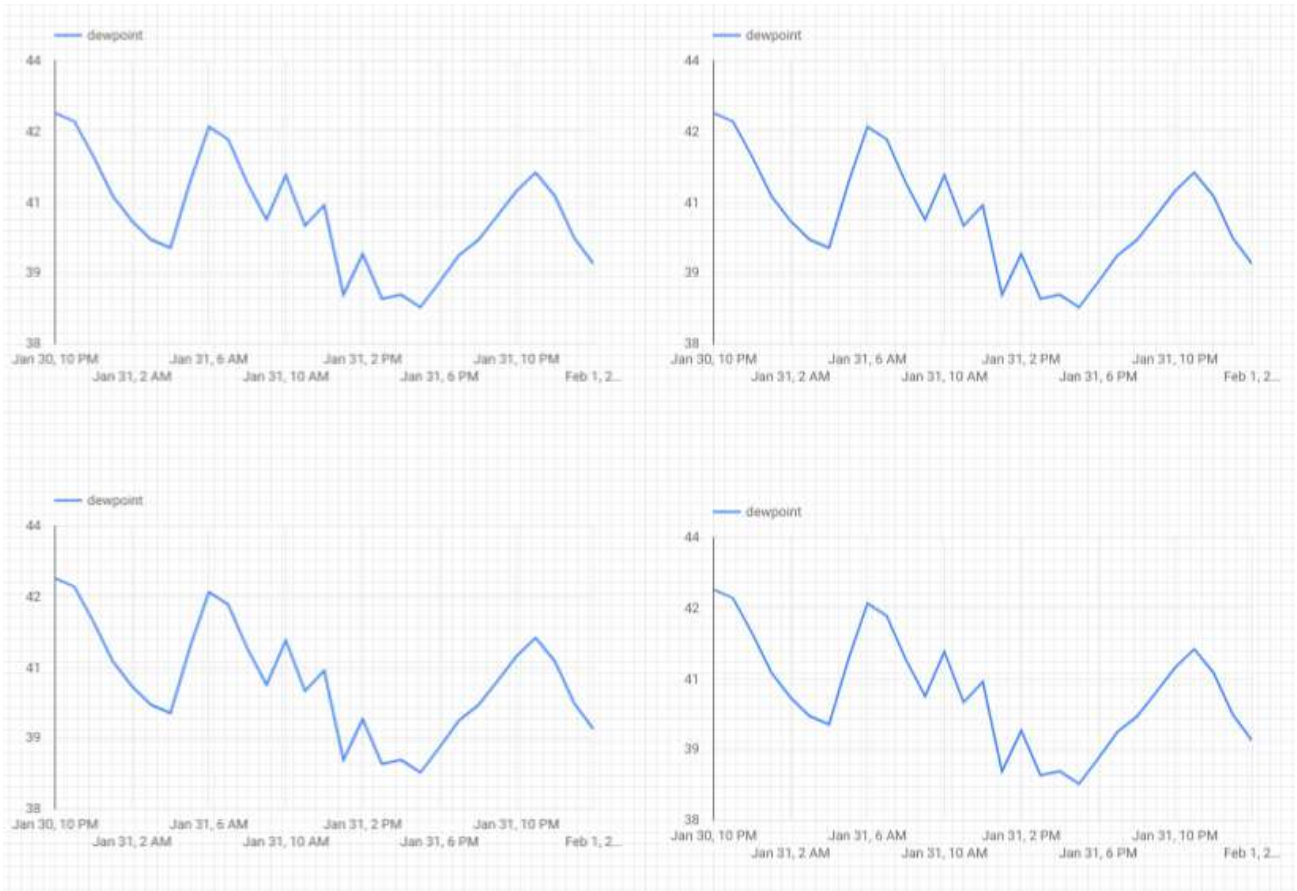
Axis Max

(auto)

Custom Tick Interval

(auto)

Click the graph on the sheet and copy/paste (Ctrl-C/Ctrl-V) it 3 times. Align the graphs so that each has $\frac{1}{4}$ of the layout




Click on each graph and under the Time Series Properties and Data section click on the existing metric (dewpoint), choose a different metric to be displayed until all four weather readings (dewpoint, temperature, humidity and pressure) have their own graph.


Time series Properties


DATA


STYLE


Chart Type


































Data Source



weatherDataTable


Dimension

Time Dimension



timecollected

Breakdown Dimension




ADD A BREAKDOWN DIMEN...

Metric

123

dewpoint




ADD A METRIC

dewpoint

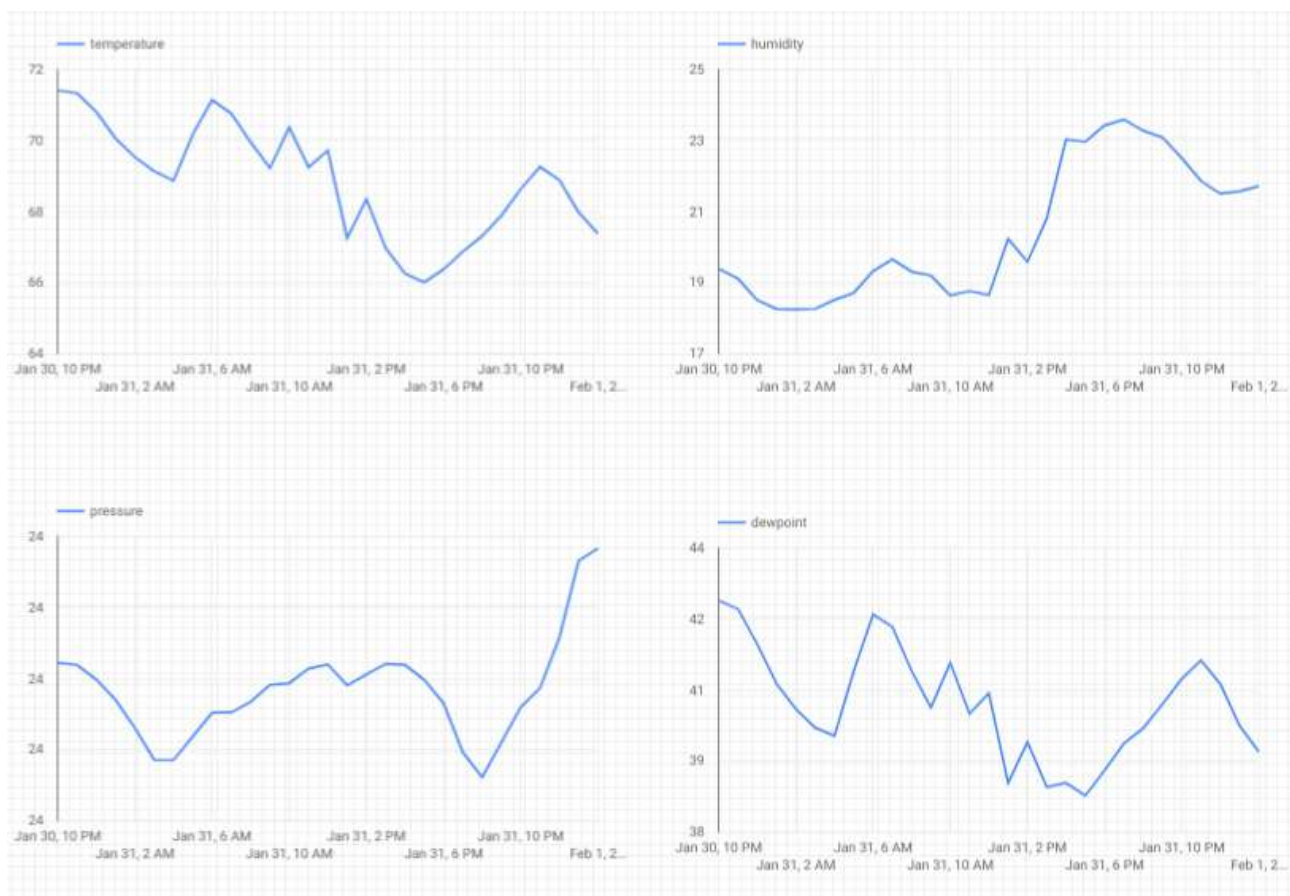
Filter

Time Series Filter



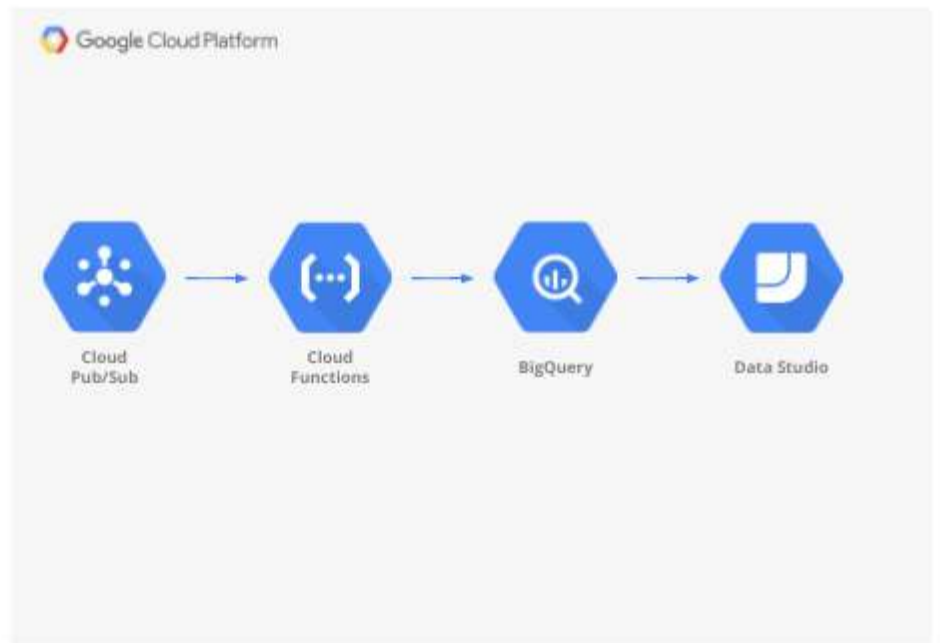
ADD A FILTER

You now have a basic dashboard!



Final output

We have created an entire data pipeline! In doing so, you've learned how to use Google Pub/Sub, how to deploy a serverless Function, how to leverage BigQuery and how to create an analytics dashboard using Data Studio. In addition, you've seen how the Google Cloud SDK can be used securely to bring data into the Google Cloud Platform. Finally, you now have some hands-on experience with an important architectural pattern that can handle high volumes while maintaining availability.



Clean-up

Once you are done experimenting with the weather data and the analytics pipeline, you can remove the running resources.

If you built the IoT sensor, shut it down. Hit Ctrl-C in the terminal window to stop the script and then type the following to power down the Raspberry Pi

```
shutdown -h now
```

Go to Cloud Functions, click on the checkbox next to function-weatherPubSubToBQ and then click on Delete

Cloud Functions

Overview

[+ CREATE FUNCTION](#)

[REFRESH](#)

[DELETE](#)


[COPY](#)


Filter functions


Columns

<input checked="" type="checkbox"/>	Name ^	Region	Trigger	Memory allocated	Executed function	Last deployed
<input checked="" type="checkbox"/>	function-weatherPubSubToBQ	us-central1	topic: weatherdata	256 MB	subscribe	1/29/18, 3:34 PM

Go to Pub/Sub, click on Topic, click on the checkbox next to the weatherdata topic and then click on Delete

 Pub/Sub

 Topics


 Subscriptions

Topics


Filter by topic name


☒ Topic name


Subscriptions


<input checked="" type="checkbox"/> projects/iot2analytics/topics/weatherdata	1	
---	---	---


Go to Storage, click on the checkboxes next to the storage buckets and then click on Delete

 Storage

 Browser

 Transfer





 Transfer Appliance

 Settings

Browser

Filter by prefix...

Columns

Name	Default storage class	Location	Lifecycle	Labels	Requester pays	
<input checked="" type="checkbox"/> lot2analytics-tmp	Multi-Regional	US	None		 Off	
<input checked="" type="checkbox"/> keystore-lot2analytics	Multi-Regional	US	None		 Off	

Go to bigquery.cloud.google.com, click the down arrow next to your project name, click the down arrow to the right of the weatherData dataset and then click on Delete dataset.

COMPOSE QUERY

Query History

Job History

Filter by ID or label

lot2Analytics

weatherData

Public Datasets

bigquery-public-data:hacker_news

bigquery-public-data:noaa_gsod

Dataset Details: weatherData

Description

Describe this dataset...

Details

Default Table Expiration

Never

Edit

Create new table

Share dataset

Delete dataset

When prompted, type in the dataset ID (weatherData) in order to finish deleting the data.



Delete Dataset

Are you sure you want to delete dataset weatherData and all of its contents?

This will delete 1 table including weatherDataTable.

This action cannot be reversed.

To confirm, type the dataset ID: weatherData

OK

Cancel