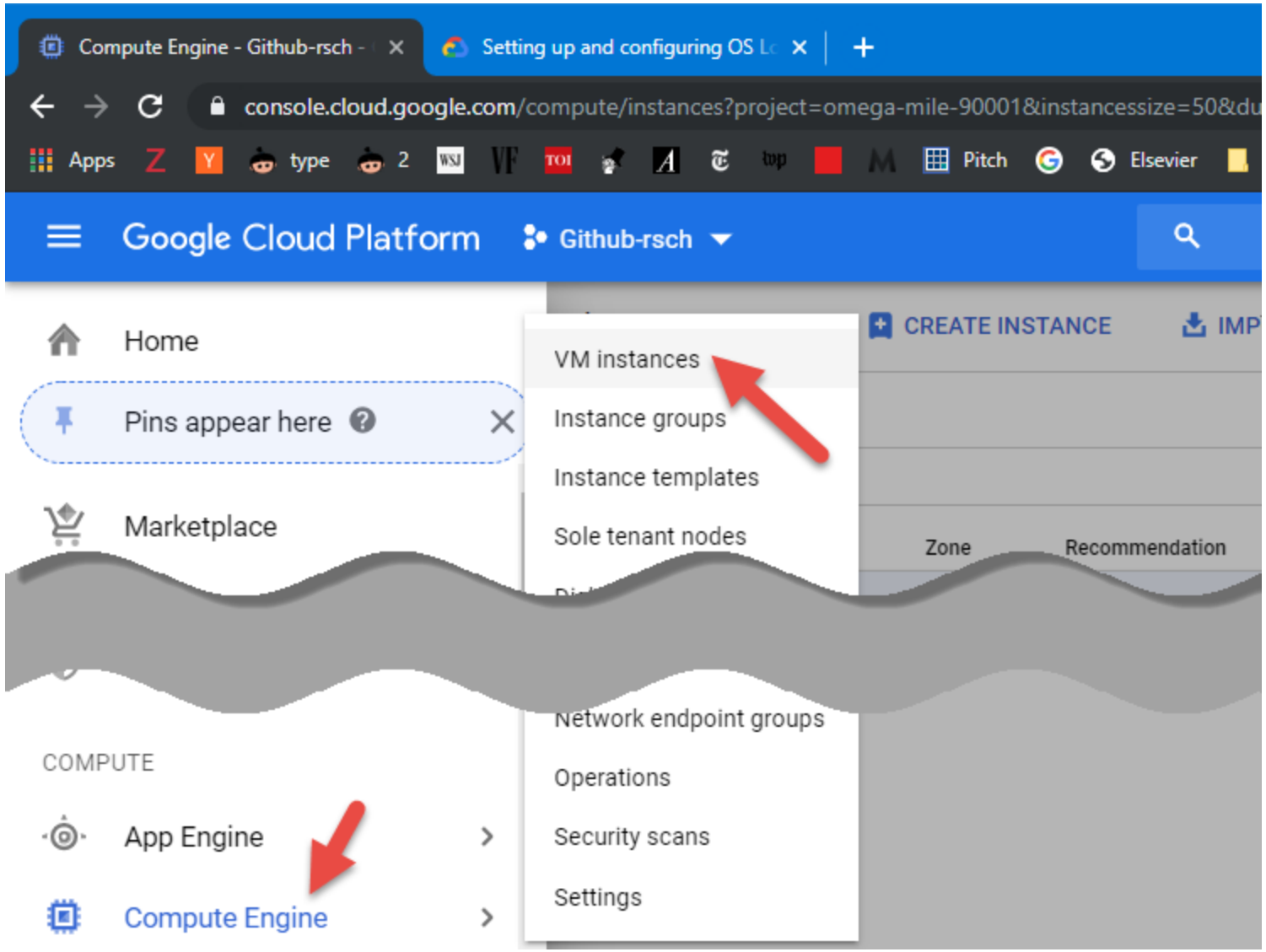
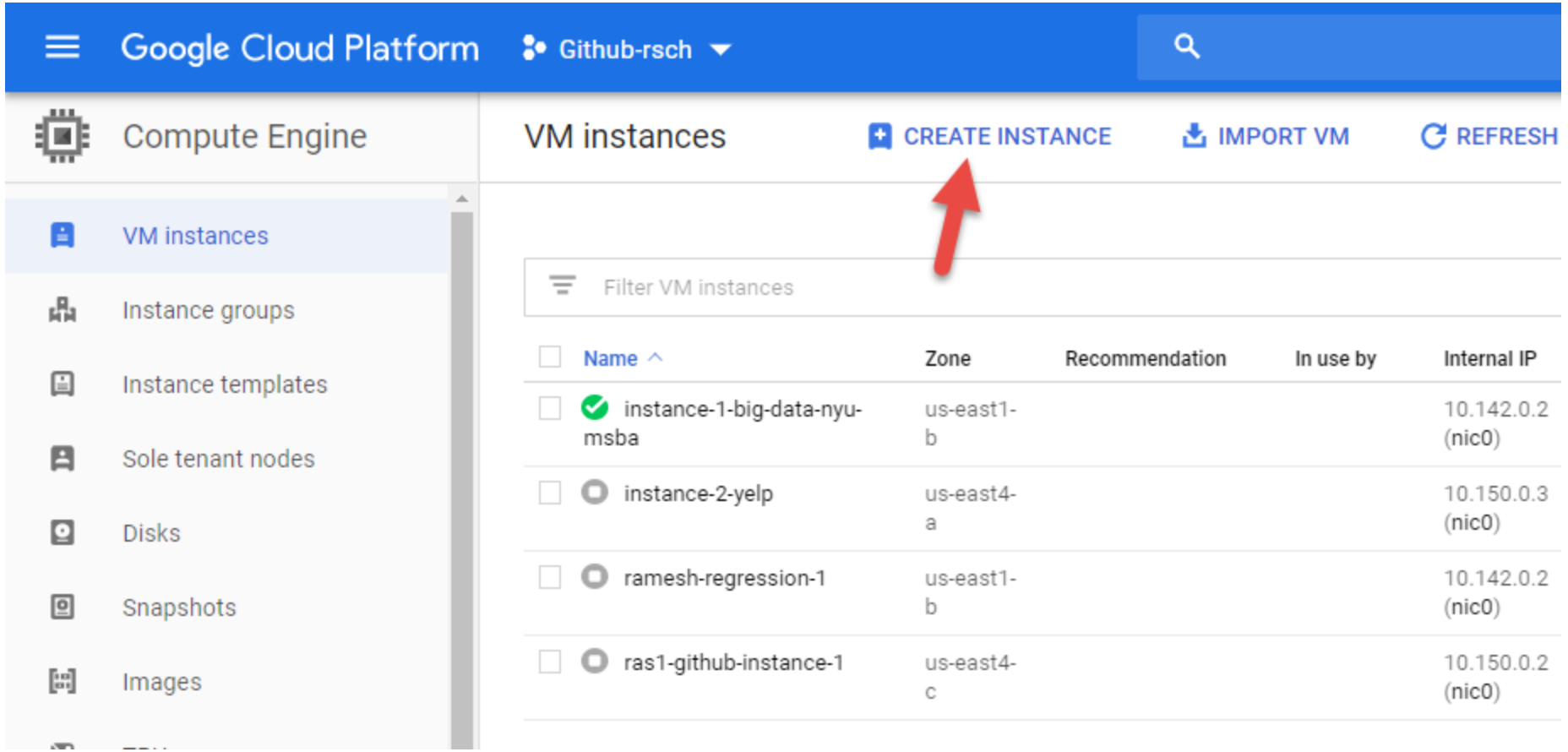
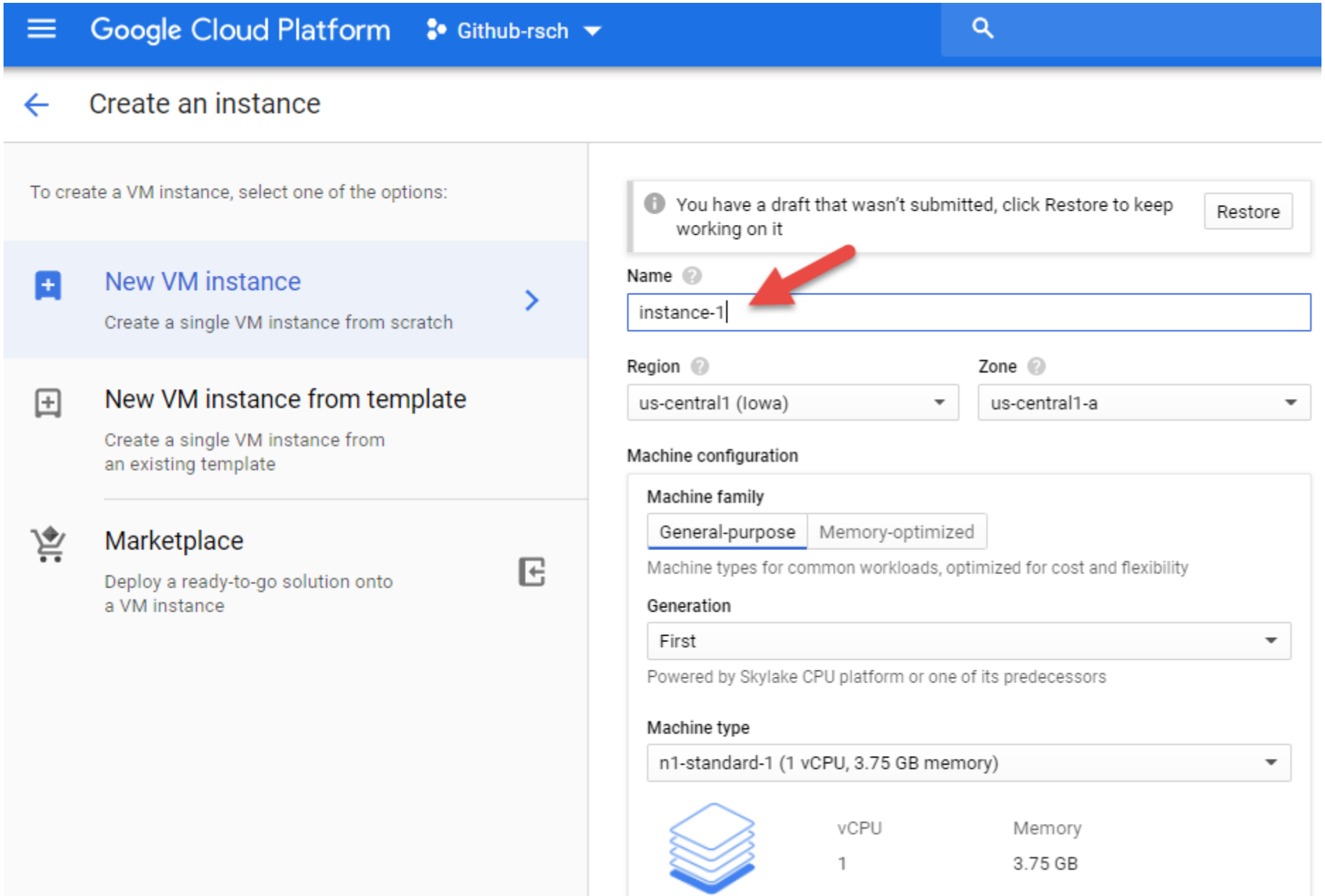
Google Cloud

Ramesh Shankar

University of Connecticut


Compute Engine - Github-rsch -
Setting up and configuring OS L
console.cloud.google.com/compute/instances?project=omega-mile-90001&instancessize=50&du
Apps
type
2
Pitch
Elsevier
Google Cloud Platform
Github-rsch
CREATE INSTANCE
Home
Pins appear here
Marketplace
VM instances
Instance groups
Instance templates
Sole tenant nodes
Zone
Recommendation
Network endpoint groups
Operations
Security scans
Settings
COMPUTE
App Engine
Compute Engine


Google Cloud Platform
Github-rsch
Compute Engine
VM instances
CREATE INSTANCE
IMPORT VM
REFRESH
VM instances
Instance groups
Instance templates
Sole tenant nodes
Disks
Snapshots
Images
Filter VM instances
Name
Zone
Recommendation
In use by
Internal IP
instance-1-big-data-nyu-msba
us-east1-b
10.142.0.2 (nic0)
instance-2-yelp
us-east4-a
10.150.0.3 (nic0)
ramesh-regression-1
us-east1-b
10.142.0.2 (nic0)
ras1-github-instance-1
us-east4-c
10.150.0.2 (nic0)


Google Cloud Platform
Github-rsch
Create an instance
To create a VM instance, select one of the options:
New VM instance
Create a single VM instance from scratch
New VM instance from template
Create a single VM instance from an existing template
Marketplace
Deploy a ready-to-go solution onto a VM instance
You have a draft that wasn't submitted, click Restore to keep working on it
Restore
Name
instance-1
Region
us-central1 (Iowa)
Zone
us-central1-a
Machine configuration
Machine family
General-purpose
Memory-optimized
Machine types for common workloads, optimized for cost and flexibility
Generation
First
Powered by Skylake CPU platform or one of its predecessors
Machine type
n1-standard-1 (1 vCPU, 3.75 GB memory)
vCPU
1
Memory
3.75 GB

Container

Deploy a container image to this VM instance. Learn more

Boot disk

New 10 GB standard persistent disk

Image

Debian GNU/Linux 9 (stretch)

Change

Identity and API access

Service account

Compute Engine default service account

Access scopes

- Allow default access
- Allow full access to all Cloud APIs
- Set access for each API

Firewall

Add tags and firewall rules to allow specific network traffic from the Internet

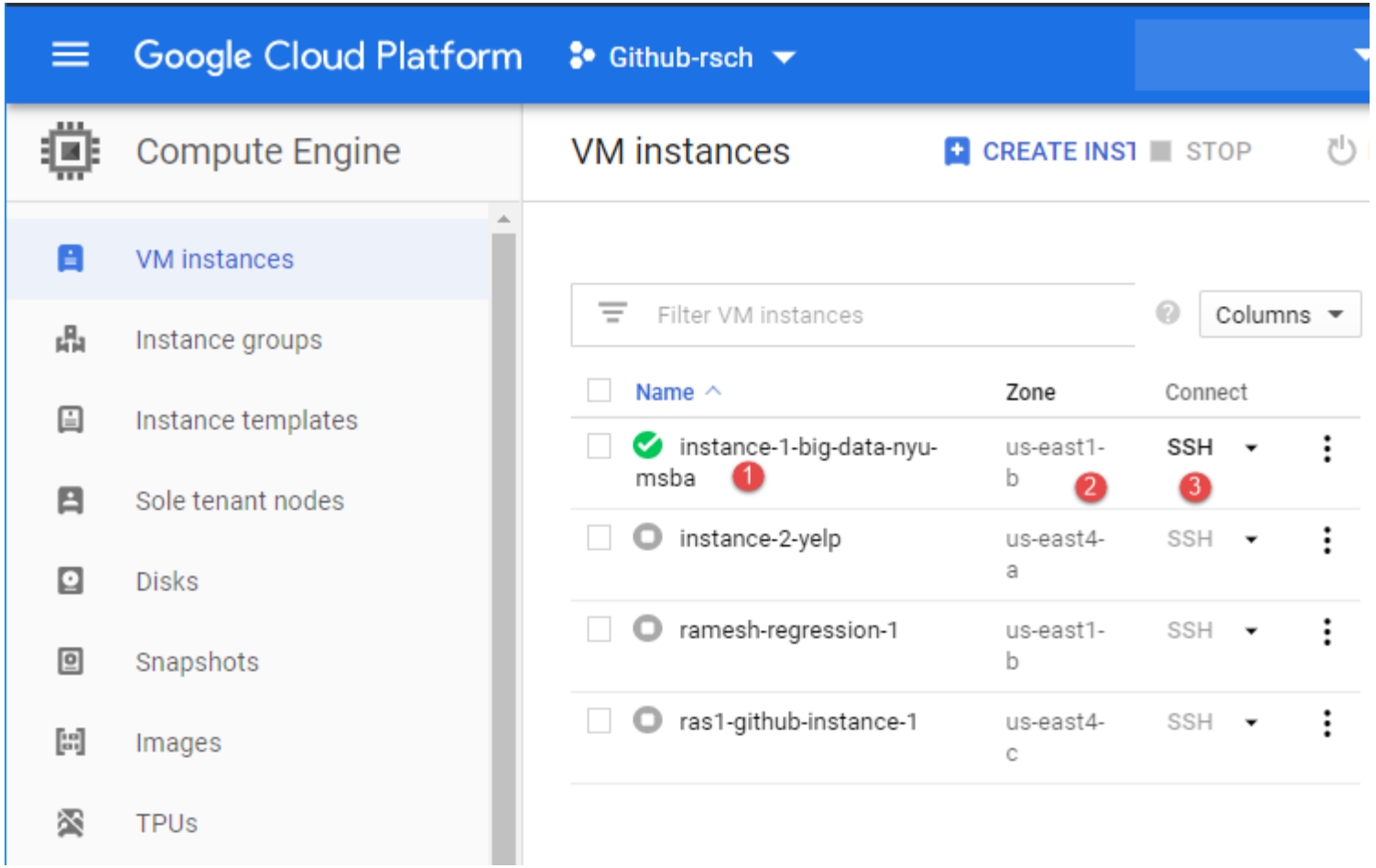
- Allow HTTP traffic
- Allow HTTPS traffic

Management, security, disks, networking, sole tenancy

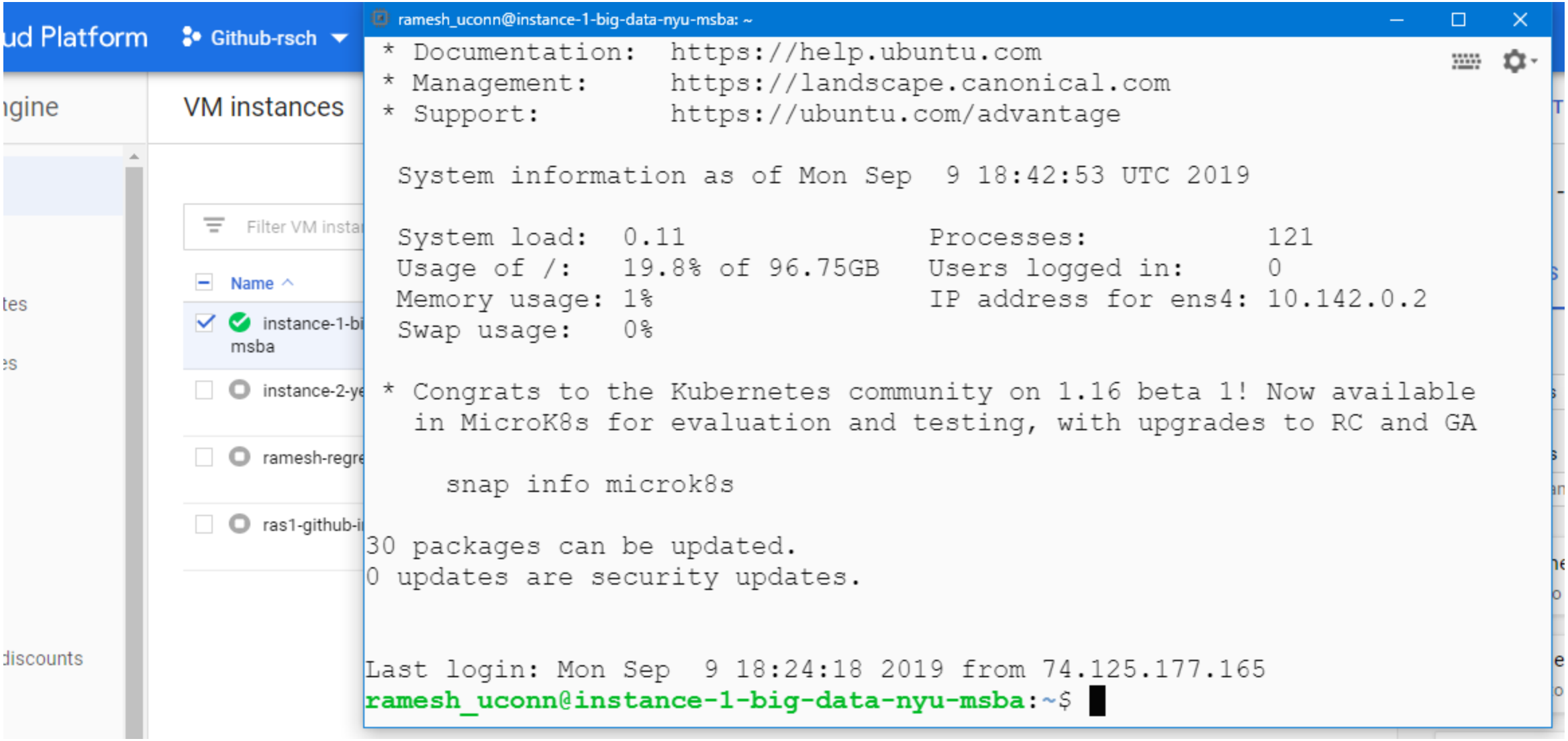
You will be billed for this instance. Compute Engine pricing

Create Cancel

Equivalent REST or command line



1. Note the instance name
2. Note the zone
3. Click on the SSH button to start your instance


ud Platform
Github-rsch
VM instances
Name
instance-1-bi
msba
instance-2-ye
ramesh-regre
ras1-github-i
ramesh_uconn@instance-1-big-data-nyu-msba: ~
* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage
System information as of Mon Sep 9 18:42:53 UTC 2019
System load: 0.11 Processes: 121
Usage of /: 19.8% of 96.75GB Users logged in: 0
Memory usage: 1% IP address for ens4: 10.142.0.2
Swap usage: 0%
* Congrats to the Kubernetes community on 1.16 beta 1! Now available
in MicroK8s for evaluation and testing, with upgrades to RC and GA
snap info microk8s
30 packages can be updated.
0 updates are security updates.
Last login: Mon Sep 9 18:24:18 2019 from 74.125.177.165
ramesh_uconn@instance-1-big-data-nyu-msba:~$

```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ more yellowtaxi_2017.sh
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-01.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-02.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-03.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-04.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-05.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-06.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-07.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-08.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-09.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-10.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-11.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-12.csv
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$
```

```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ more yellowtaxi_2018.sh
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-01.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-02.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-03.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-04.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-05.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-06.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-07.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-08.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-09.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-10.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-11.csv
wget https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-12.csv
```

```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ sh yellowtaxi_2017.sh
--2019-09-09 18:55:15--  https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-01.csv
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.97.181
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.97.181|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 854903002 (815M) [application/x-www-form-urlencoded]
Saving to: 'yellow_tripdata_2017-01.csv'

yellow_tripdata_2017-01.c 100%[===================================>] 815.30M  48.9MB/s    in 15s

2019-09-09 18:55:31 (54.1 MB/s) - 'yellow_tripdata_2017-01.csv' saved [854903002/854903002]

--2019-09-09 18:55:31--  https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2017-02.csv
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.176.141
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.176.141|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 808065449 (771M) [application/x-www-form-urlencoded]
Saving to: 'yellow_tripdata_2017-02.csv'

yellow_tripdata_2017-02.c  21%[=======>                            ] 163.48M  59.7MB/s
```

1

2

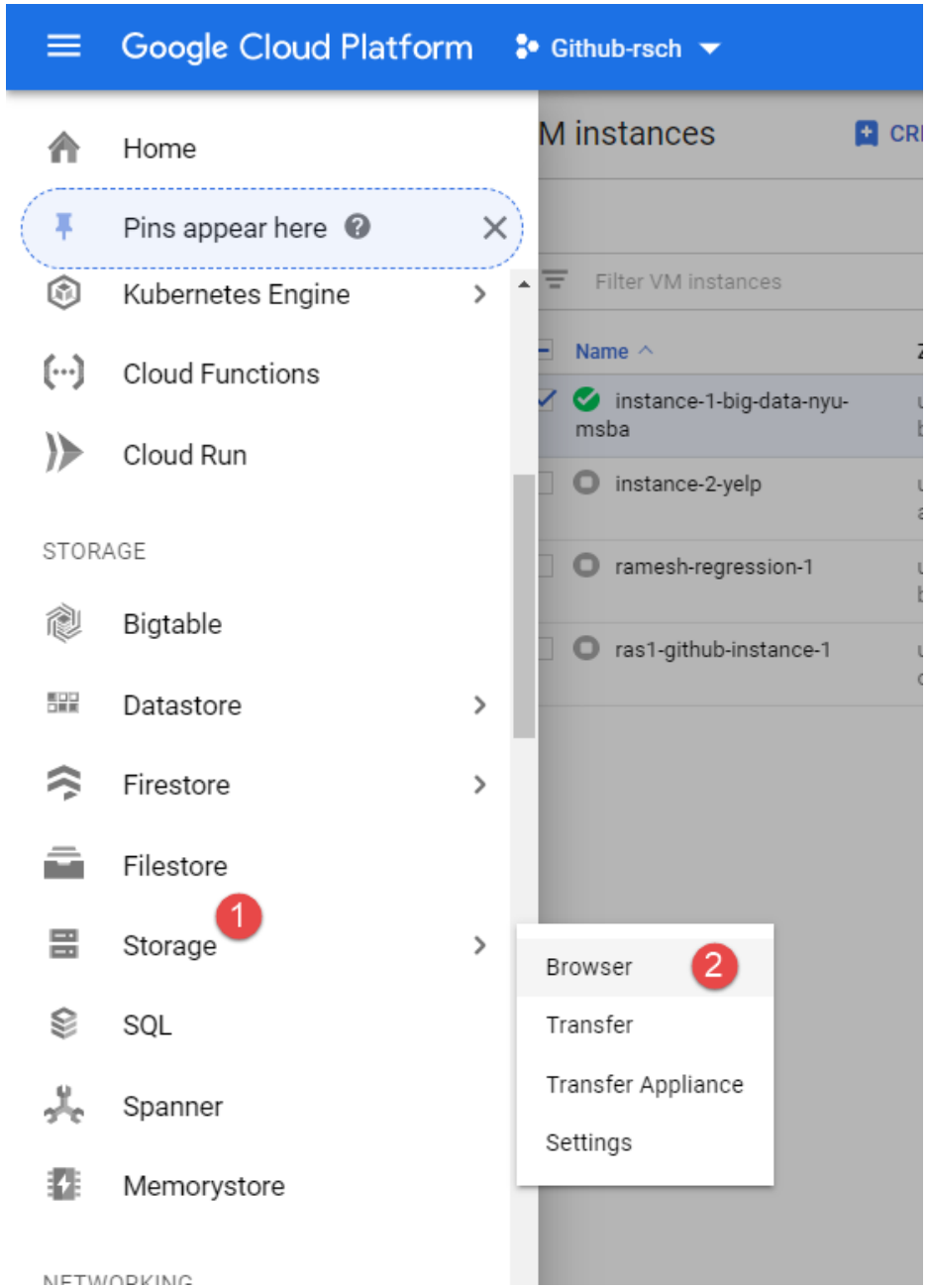
3

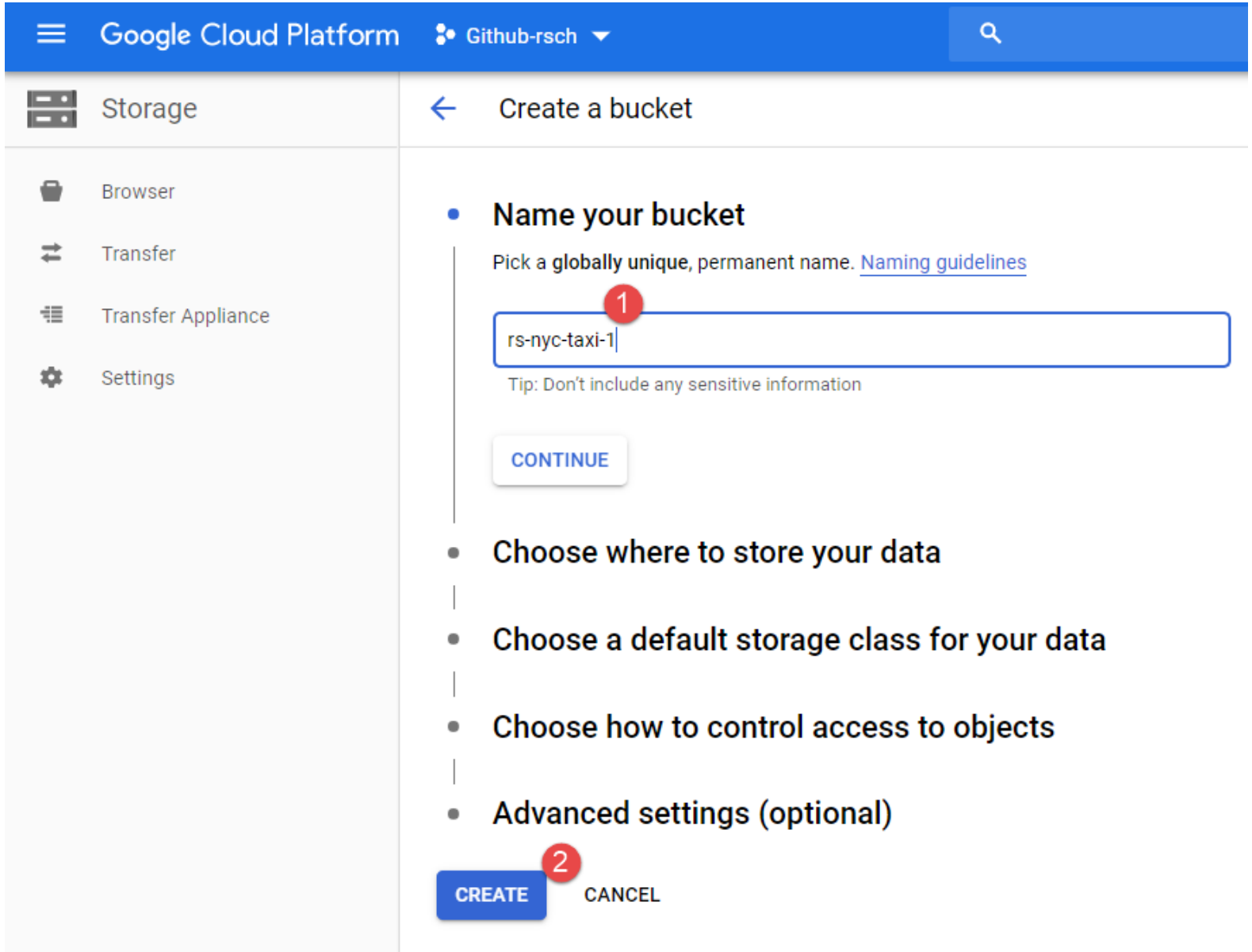
```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ sh yellowtaxi_2018.sh
--2019-09-09 18:58:29--  https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-01.csv
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.105.141
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.105.141|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 772098307 (736M) [text/csv]
Saving to: 'yellow_tripdata_2018-01.csv'

yellow_tripdata_2018-01.c 100%[===================================>] 736.33M  69.3MB/s    in 8.9s

2019-09-09 18:58:38 (82.9 MB/s) - 'yellow_tripdata_2018-01.csv' saved [772098307/772098307]

--2019-09-09 18:58:38--  https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2018-02.csv
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.105.141
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.105.141|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 748827487 (714M) [text/csv]
Saving to: 'yellow_tripdata_2018-02.csv'

yellow_tripdata_2018-02.c  29%[==========>                          ] 207.99M  90.4MB/s
```

1

2

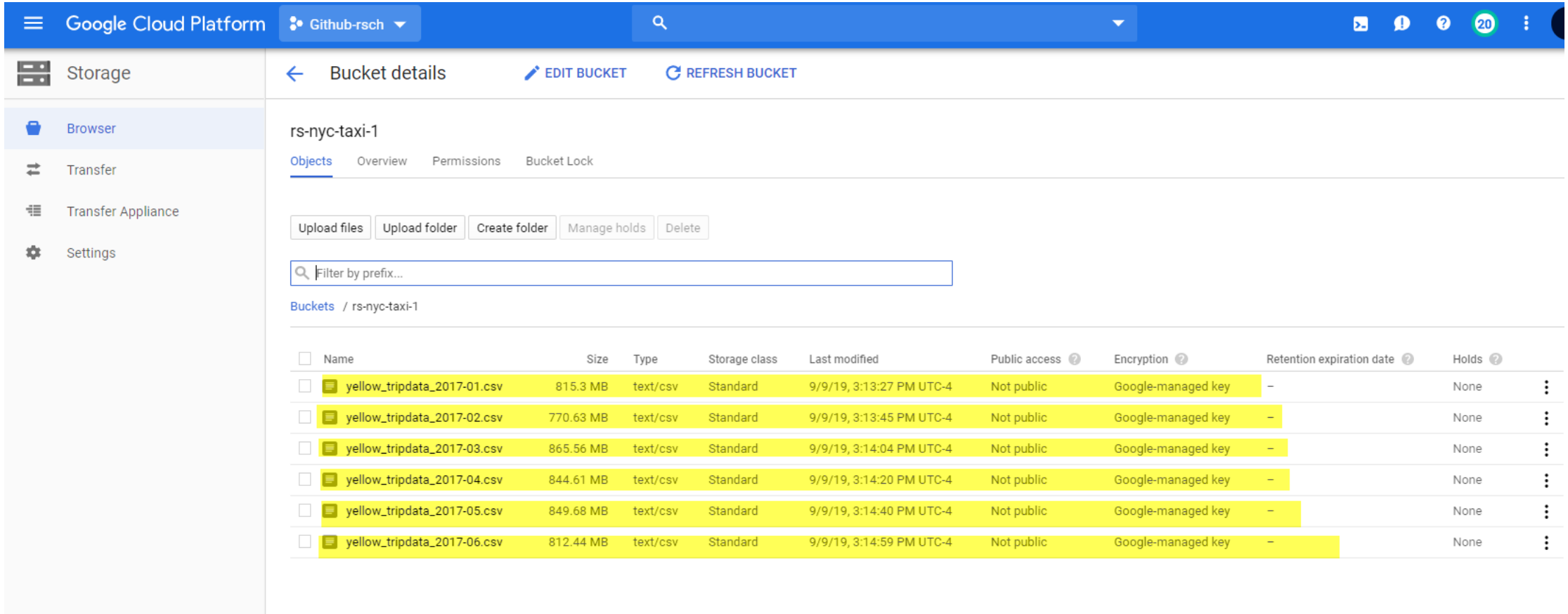
3

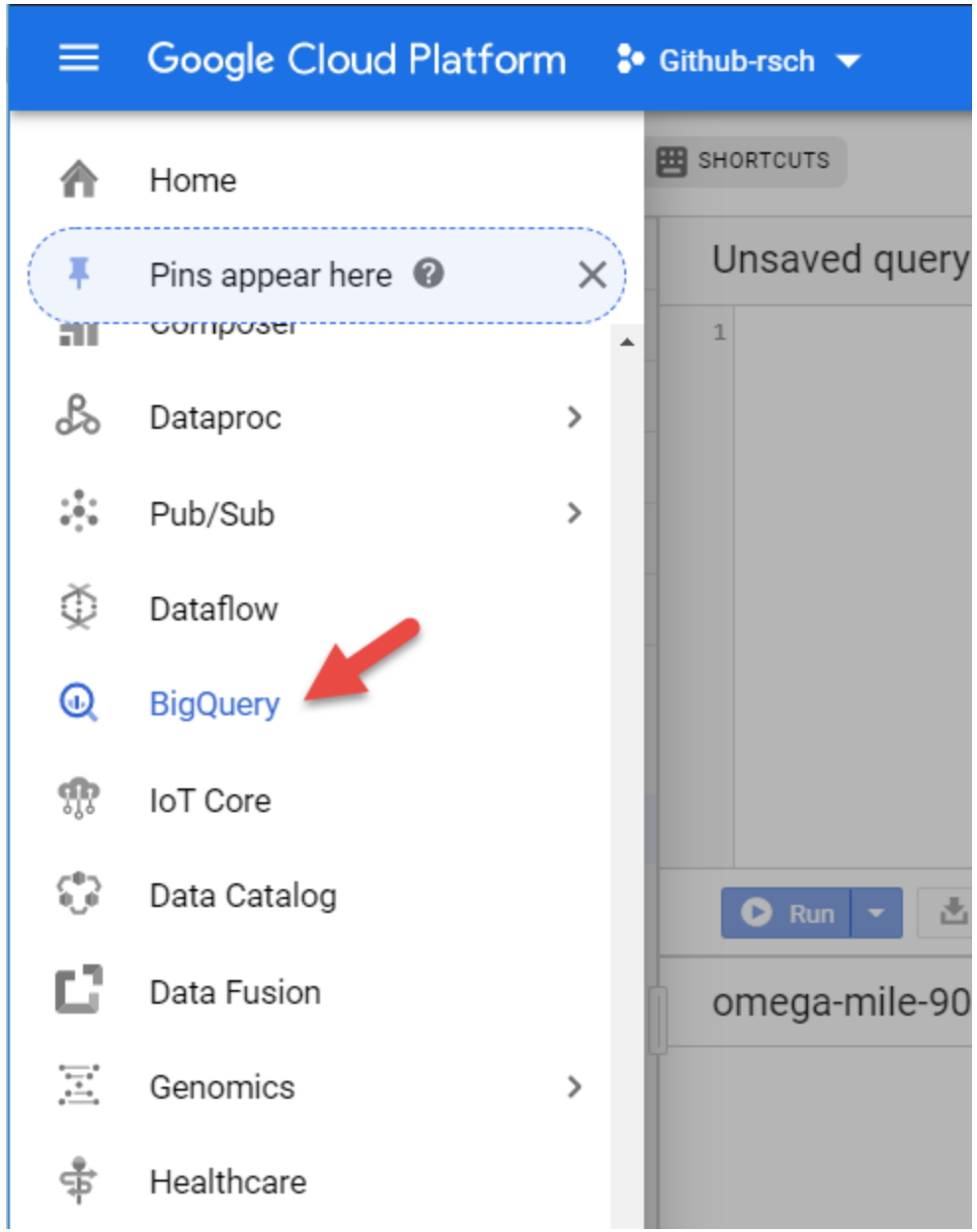

Google Cloud Platform
Github-rsch
Home
Pins appear here
Kubernetes Engine
Cloud Functions
Cloud Run
STORAGE
Bigtable
Datastore
Firestore
Filestore
Storage
1
SQL
Spanner
Memorystore
M instances
Filter VM instances
Name
instance-1-big-data-nyu-msba
instance-2-yelp
ramesh-regression-1
ras1-github-instance-1
Browser
2
Transfer
Transfer Appliance
Settings

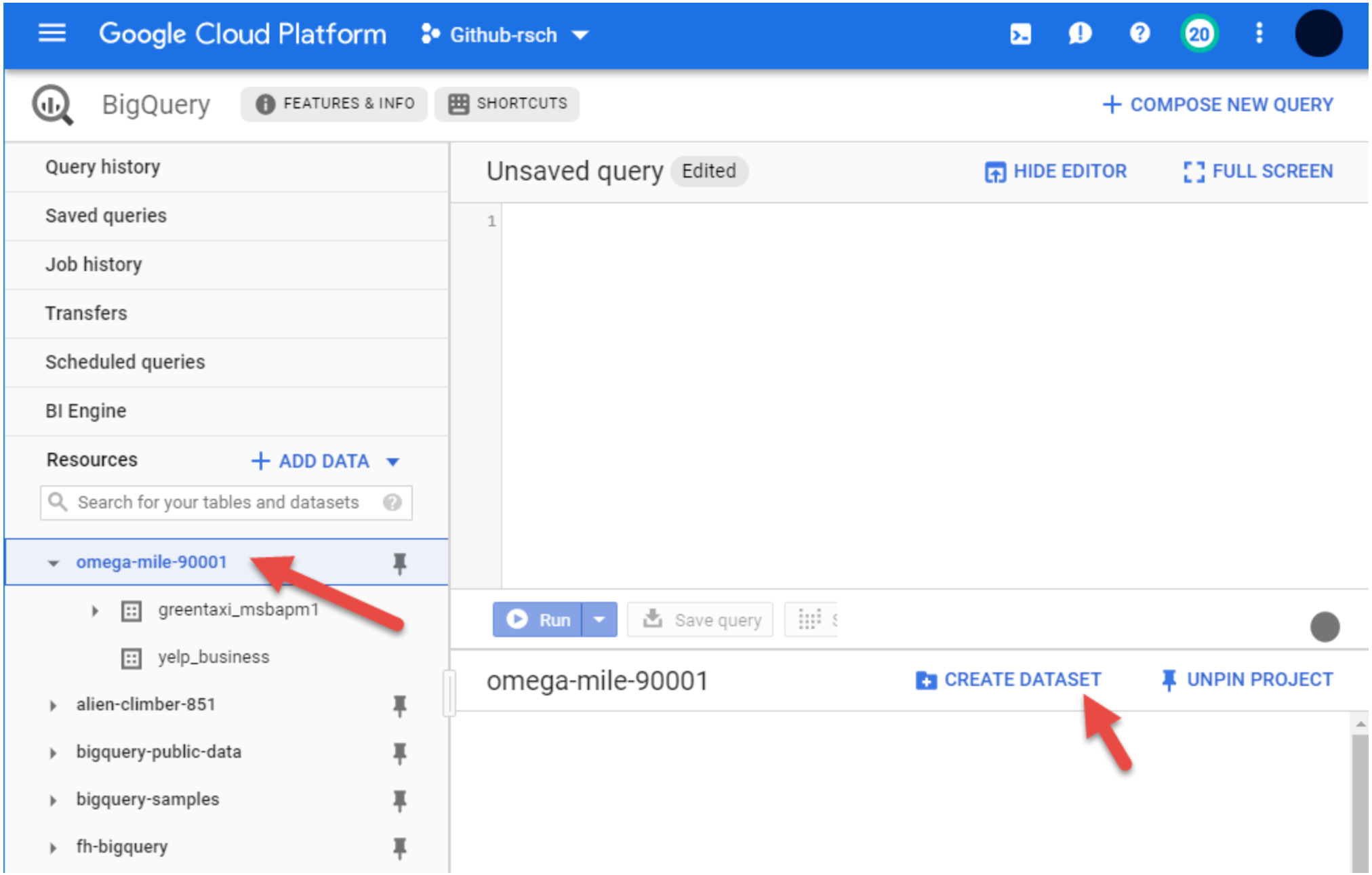

Google Cloud Platform
Github-rsch
Storage
Browser
Transfer
Transfer Appliance
Settings
Create a bucket
Name your bucket
Pick a globally unique, permanent name. Naming guidelines
1
rs-nyc-taxi-1
Tip: Don't include any sensitive information
CONTINUE
Choose where to store your data
Choose a default storage class for your data
Choose how to control access to objects
Advanced settings (optional)
2
CREATE
CANCEL

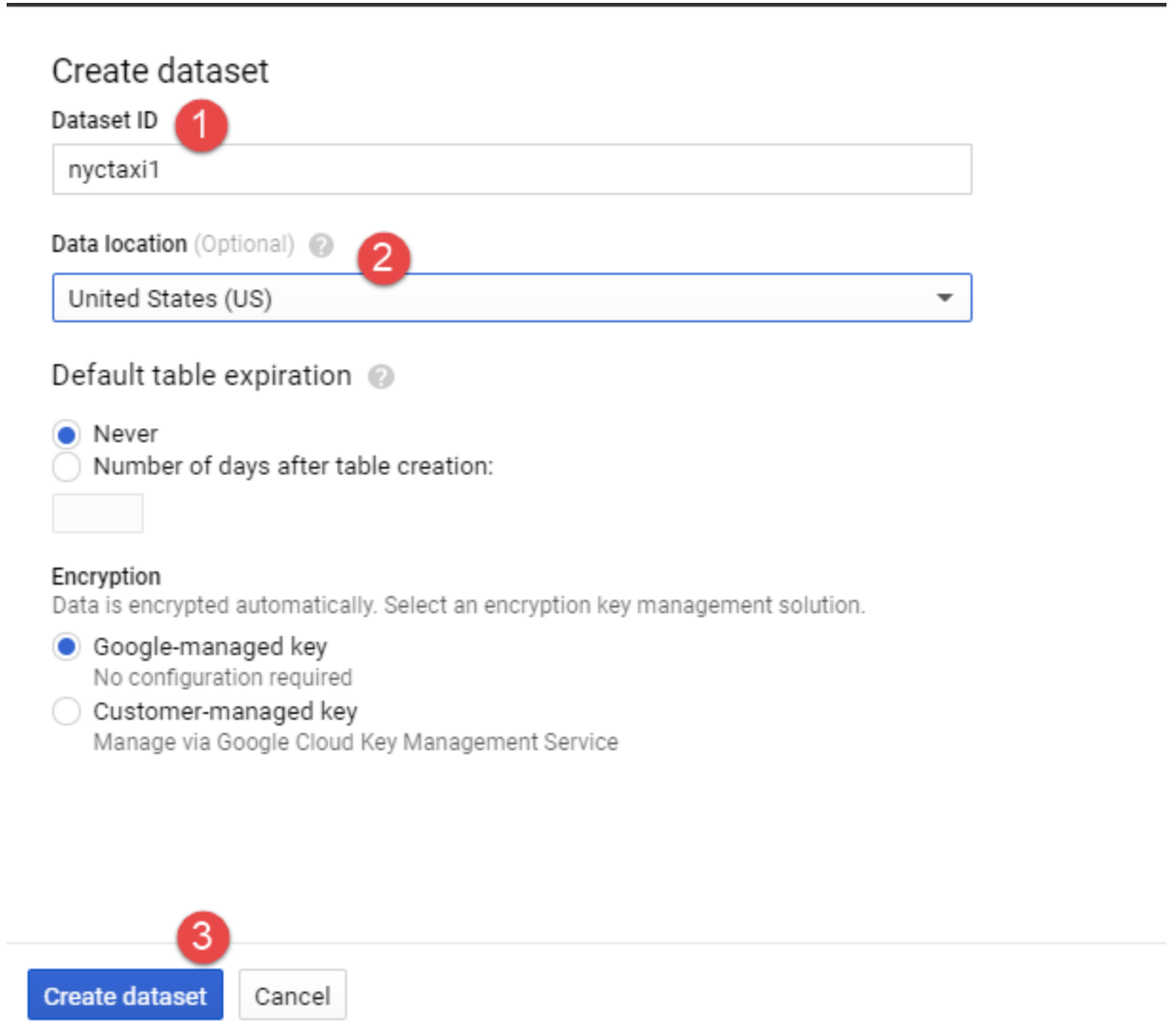
```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ more yellowtaxi_2017_2018_upload.sh
gsutil cp yellow_tripdata_2017-01.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-02.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-03.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-04.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-05.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-06.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-07.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-08.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-09.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-10.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-11.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2017-12.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-01.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-02.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-03.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-04.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-05.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-06.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-07.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-08.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-09.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-10.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow_tripdata_2018-11.csv us-east1 gs://rs-nyc-taxi-1
gsutil cp yellow tripdata 2018-12.csv us-east1 gs://rs-nyc-taxi-1
```

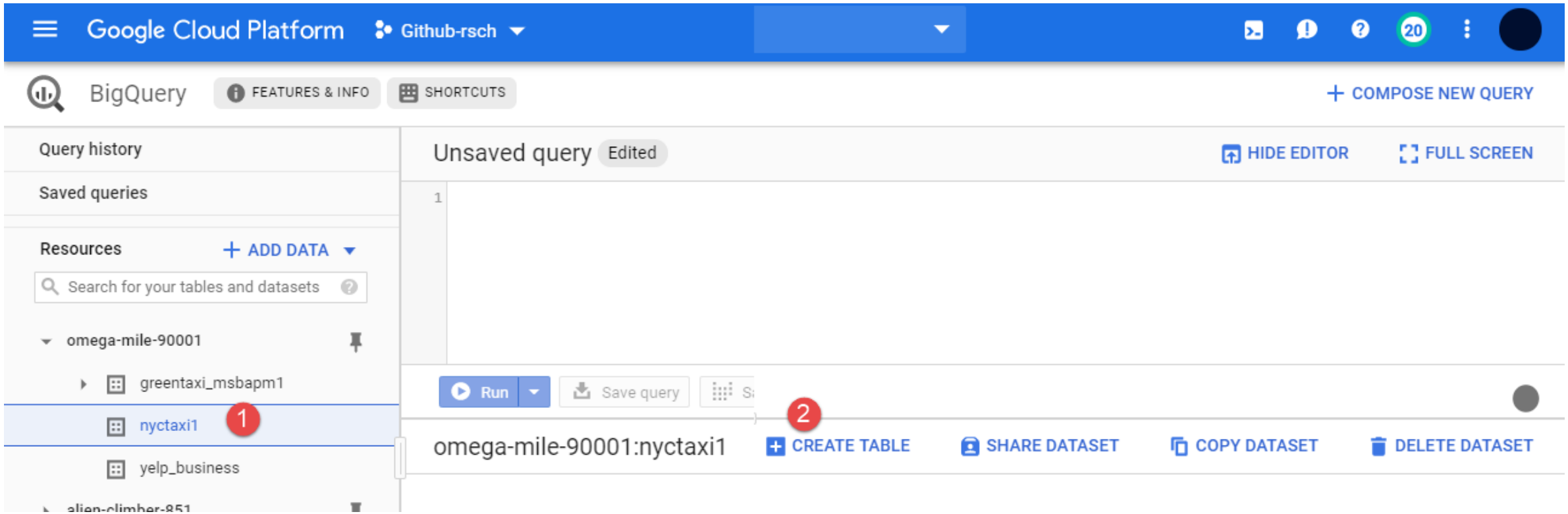
```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ sh yellowtaxi_2017_2018_upload.sh
Copying file://yellow_tripdata_2017-01.csv [Content-Type=text/csv]...
==> NOTE: You are uploading one or more large file(s), which would run
significantly faster if you enable parallel composite uploads. This
feature can be enabled by editing the
"parallel_composite_upload_threshold" value in your .boto
configuration file. However, note that if you do this large files will
be uploaded as `composite objects
<https://cloud.google.com/storage/docs/composite-objects>`_,which
means that any user who downloads such objects will need to have a
compiled crcmod installed (see "gsutil help crcmod"). This is because
without a compiled crcmod, computing checksums on composite objects is
so slow that gsutil disables downloads of composite objects.

CommandException: No URLs matched: us-east1/s
Copying file://yellow_tripdata_2017-02.csv [Content-Type=text/csv]...
==> NOTE: You are uploading one or more large file(s), which would run
significantly faster if you enable parallel composite uploads. This
```

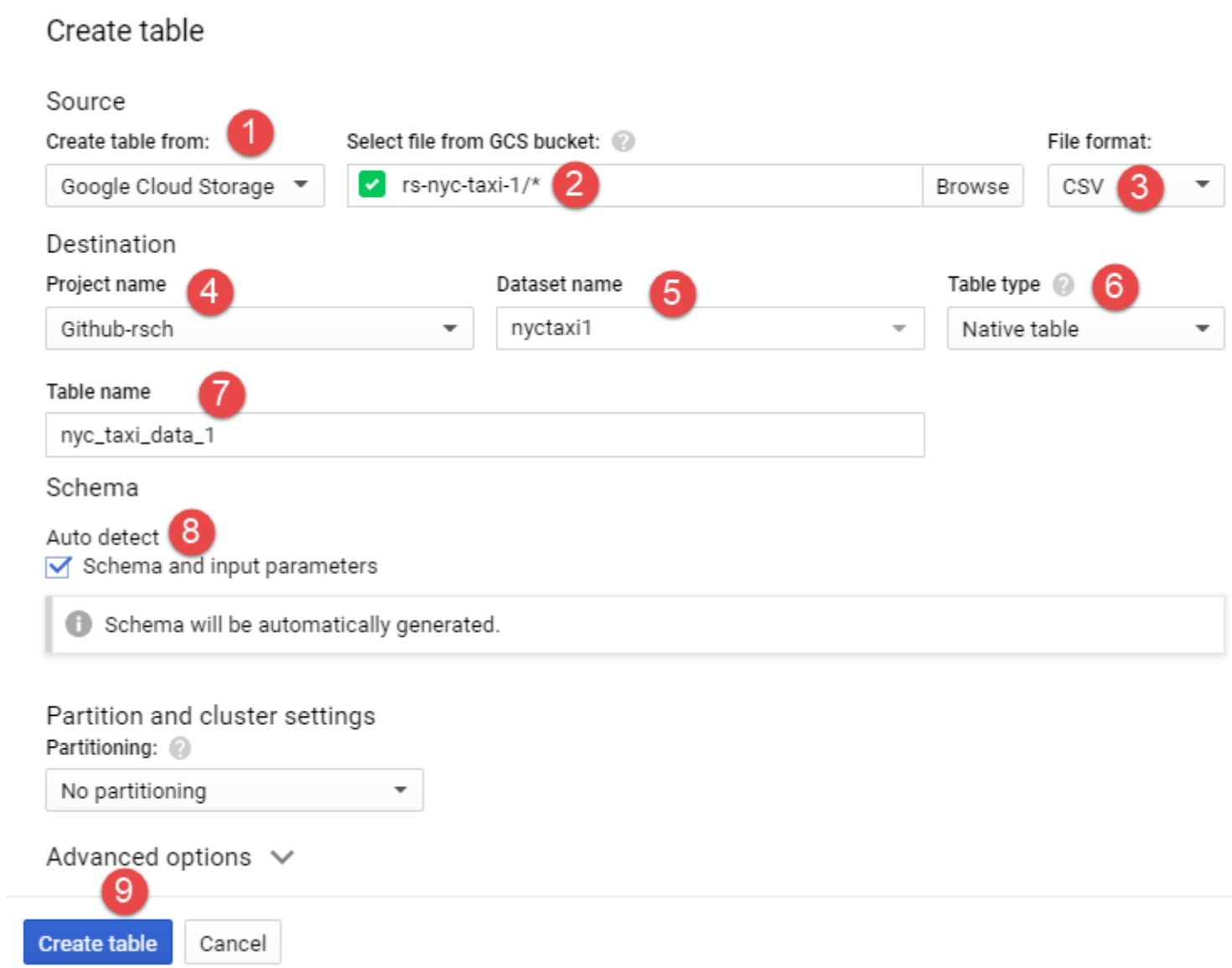

Google Cloud Platform
Github-rsch
20
Storage
Browser
Transfer
Transfer Appliance
Settings
Bucket details
EDIT BUCKET
REFRESH BUCKET
rs-nyc-taxi-1
Objects
Overview
Permissions
Bucket Lock
Upload files
Upload folder
Create folder
Manage holds
Delete
Filter by prefix...
Buckets / rs-nyc-taxi-1
Name
Size
Type
Storage class
Last modified
Public access
Encryption
Retention expiration date
Holds
yellow_tripdata_2017-01.csv 815.3 MB text/csv Standard 9/9/19, 3:13:27 PM UTC-4 Not public Google-managed key – None
yellow_tripdata_2017-02.csv 770.63 MB text/csv Standard 9/9/19, 3:13:45 PM UTC-4 Not public Google-managed key – None
yellow_tripdata_2017-03.csv 865.56 MB text/csv Standard 9/9/19, 3:14:04 PM UTC-4 Not public Google-managed key – None
yellow_tripdata_2017-04.csv 844.61 MB text/csv Standard 9/9/19, 3:14:20 PM UTC-4 Not public Google-managed key – None
yellow_tripdata_2017-05.csv 849.68 MB text/csv Standard 9/9/19, 3:14:40 PM UTC-4 Not public Google-managed key – None
yellow_tripdata_2017-06.csv 812.44 MB text/csv Standard 9/9/19, 3:14:59 PM UTC-4 Not public Google-managed key – None

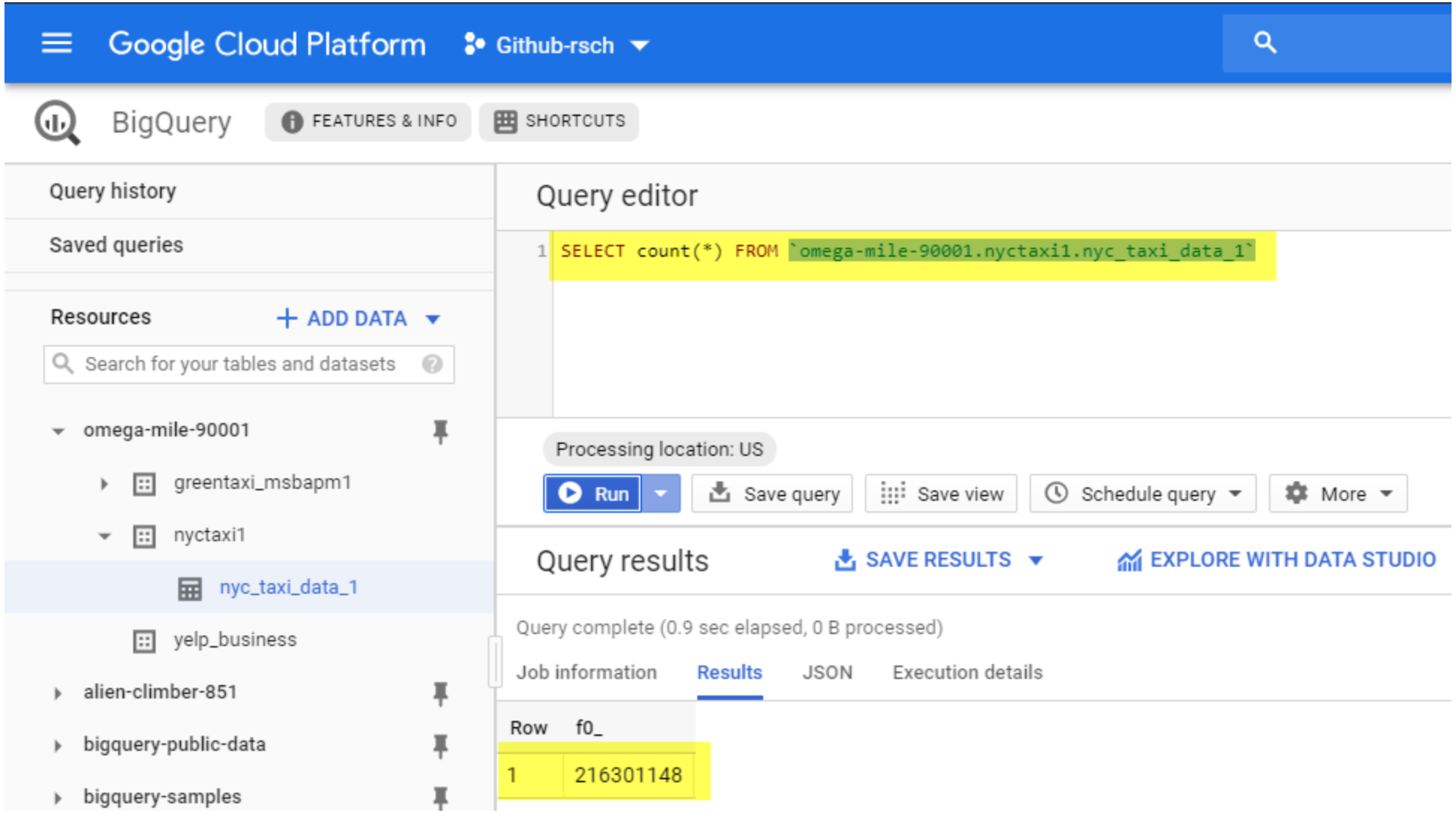

Google Cloud Platform
Github-rsch
Home
Pins appear here
Dataproc
Pub/Sub
Dataflow
BigQuery
IoT Core
Data Catalog
Data Fusion
Genomics
Healthcare
SHORTCUTS
Unsaved query
1
Run
omega-mile-90


Google Cloud Platform
Github-rsch
20
BigQuery
FEATURES & INFO
SHORTCUTS
COMPOSE NEW QUERY
Query history
Saved queries
Job history
Transfers
Scheduled queries
BI Engine
Resources
ADD DATA
Search for your tables and datasets
omega-mile-90001
greentaxi_msbapm1
yelp_business
alien-climber-851
bigquery-public-data
bigquery-samples
fh-bigquery
Unsaved query
Edited
HIDE EDITOR
FULL SCREEN
1
Run
Save query
omega-mile-90001
CREATE DATASET
UNPIN PROJECT

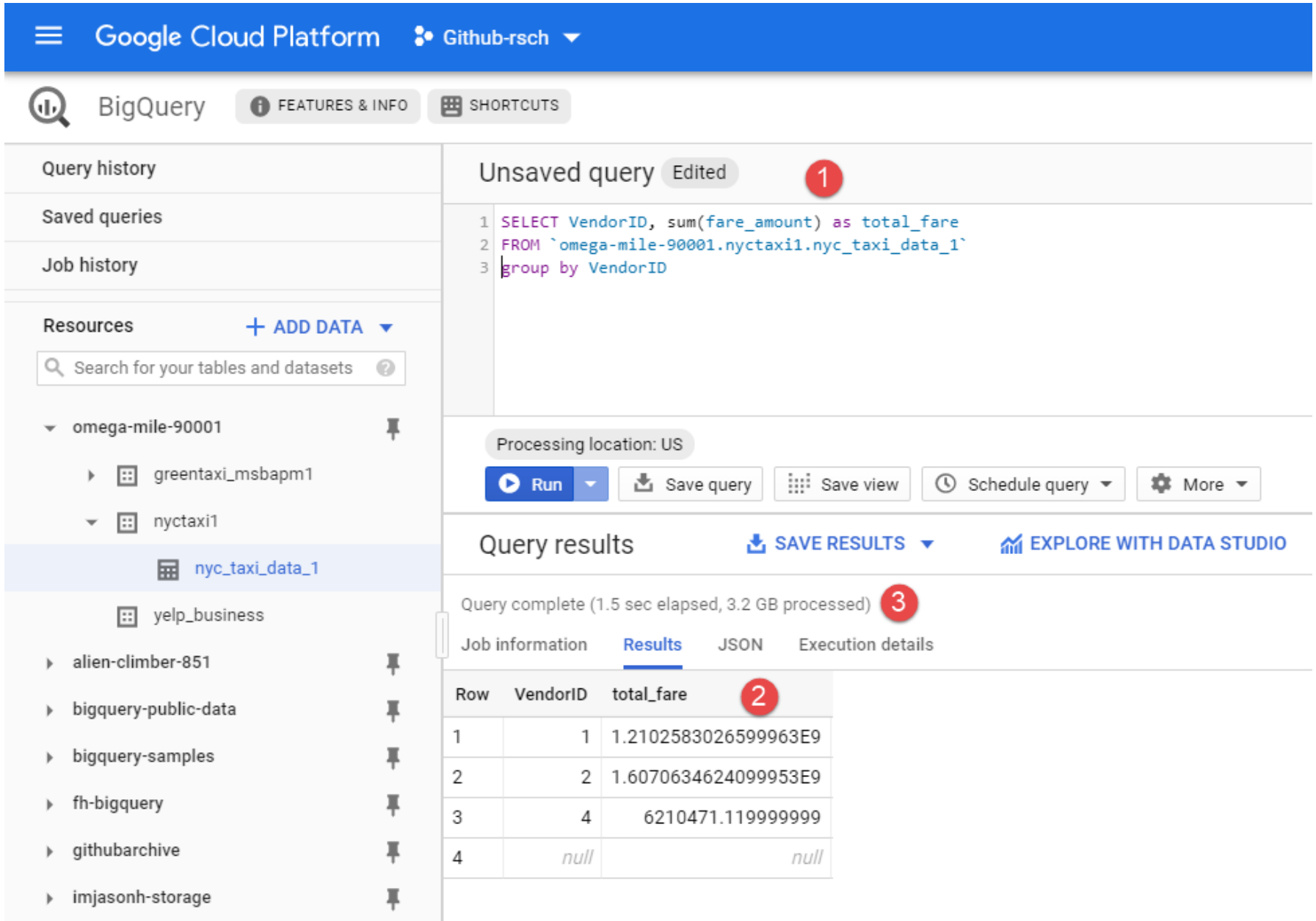

Create dataset
Dataset ID
1
nyctaxi1
Data location (Optional)
2
United States (US)
Default table expiration
Never
Number of days after table creation:
Encryption
Data is encrypted automatically. Select an encryption key management solution.
Google-managed key
No configuration required
Customer-managed key
Manage via Google Cloud Key Management Service
3
Create dataset
Cancel


Google Cloud Platform
Github-rsch
20
BigQuery
FEATURES & INFO
SHORTCUTS
COMPOSE NEW QUERY
Query history
Saved queries
Resources
ADD DATA
Search for your tables and datasets
omega-mile-90001
greentaxi_msbapm1
nyctaxi1
1
yelp_business
Unsaved query
Edited
HIDE EDITOR
FULL SCREEN
1
Run
Save query
2
omega-mile-90001:nyctaxi1
CREATE TABLE
SHARE DATASET
COPY DATASET
DELETE DATASET


Create table
Source
Create table from:
1
Google Cloud Storage
Select file from GCS bucket:
rs-nyc-taxi-1/*
2
Browse
File format:
CSV
3
Destination
Project name
4
Github-rsch
Dataset name
5
nyctaxi1
Table type
6
Native table
Table name
7
nyc_taxi_data_1
Schema
Auto detect
8
Schema and input parameters
Schema will be automatically generated.
Partition and cluster settings
Partitioning:
No partitioning
Advanced options
9
Create table
Cancel


Google Cloud Platform
Github-rsch
BigQuery
FEATURES & INFO
SHORTCUTS
Query history
Saved queries
Resources
ADD DATA
Search for your tables and datasets
omega-mile-90001
greentaxi_msbapm1
nyctaxi1
nyc_taxi_data_1
yelp_business
alien-climber-851
bigquery-public-data
bigquery-samples
Query editor
1 SELECT count(*) FROM `omega-mile-90001.nyctaxi1.nyc_taxi_data_1`
Processing location: US
Run
Save query
Save view
Schedule query
More
Query results
SAVE RESULTS
EXPLORE WITH DATA STUDIO
Query complete (0.9 sec elapsed, 0 B processed)
Job information
Results
JSON
Execution details
Row
f0_
1
216301148

```
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ wc -l yellow_tripdata_2018-11.csv
8145166 yellow_tripdata_2018-11.csv
ramesh_uconn@instance-1-big-data-nyu-msba:~/taxi$ wc -l *.csv
   9710126 yellow_tripdata_2017-01.csv
   9168827 yellow_tripdata_2017-02.csv
  10294630 yellow_tripdata_2017-03.csv
  10046190 yellow_tripdata_2017-04.csv
  10102126 yellow_tripdata_2017-05.csv
   9656995 yellow_tripdata_2017-06.csv
   8588488 yellow_tripdata_2017-07.csv
   8422155 yellow_tripdata_2017-08.csv
   8945461 yellow_tripdata_2017-09.csv
   9768817 yellow_tripdata_2017-10.csv
   9284805 yellow_tripdata_2017-11.csv
   9508278 yellow_tripdata_2017-12.csv
   8759876 yellow_tripdata_2018-01.csv
   8492078 yellow_tripdata_2018-02.csv
   9430378 yellow_tripdata_2018-03.csv
   9305517 yellow_tripdata_2018-04.csv
   9224065 yellow_tripdata_2018-05.csv
   8713833 yellow_tripdata_2018-06.csv
   7849750 yellow_tripdata_2018-07.csv
   7849136 yellow_tripdata_2018-08.csv
   8040135 yellow_tripdata_2018-09.csv
   8821107 yellow_tripdata_2018-10.csv
   8145166 yellow_tripdata_2018-11.csv
   8173233 yellow_tripdata_2018-12.csv
 216301172 total
```


Google Cloud Platform
Github-rsch
BigQuery
FEATURES & INFO
SHORTCUTS
Query history
Saved queries
Job history
Resources
ADD DATA
Search for your tables and datasets
omega-mile-90001
greentaxi_msbapm1
nyctaxi1
nyc_taxi_data_1
yelp_business
alien-climber-851
bigquery-public-data
bigquery-samples
fh-bigquery
githubarchive
imjasonh-storage
Unsaved query
Edited
1
SELECT VendorID, sum(fare_amount) as total_fare
FROM `omega-mile-90001.nyctaxi1.nyc_taxi_data_1`
group by VendorID
Processing location: US
Run
Save query
Save view
Schedule query
More
Query results
SAVE RESULTS
EXPLORE WITH DATA STUDIO
Query complete (1.5 sec elapsed, 3.2 GB processed)
3
Job information
Results
JSON
Execution details
Row
VendorID
total_fare
2
1
1
1.2102583026599963E9
2
2
1.6070634624099953E9
3
4
6210471.119999999
4
null
null