Cloud Storage

Landing
CSV and JSON Files

Silver
Parquet Files

BigQuery

Pyspark
transform
json/csv to
parquet

Pyspark join
and transform
data and
save to
BigQuery

Cloud
Dataproc

Cloud
Dataproc

Airflow for providing
Dataproc Environment
and executes the job

Cloud
Composer

Google Cloud Platform   Data Sprints   Search products and resources

Browser   CREATE BUCKET   DELETE   REFRESH   SHOW INFO P

Filter Filter buckets

Name ↑	Created	Location type	Location	Default storage class	Updated
dataproc-temp-us-east1-8886357930...	Aug 14, 2021, 10:16:28 AM	Region	us-east1 (Sout...	Standard	Aug 14, 2021, 10:16:28 AM
dataproc_temp_001	Aug 11, 2021, 11:45:23 PM	Region	us-east1 (Sout...	Standard	Aug 11, 2021, 11:45:23 PM
sprints-apps	Aug 12, 2021, 12:21:23 AM	Region	us-east1 (Sout...	Standard	Aug 12, 2021, 12:21:23 AM
sprints-dataproc	Aug 14, 2021, 10:16:14 AM	Region	us-east1 (Sout...	Standard	Aug 14, 2021, 10:16:14 AM
sprints_landing	Aug 11, 2021, 4:29:00 PM	Region	us-east1 (Sout...	Standard	Aug 11, 2021, 4:29:00 PM
sprints_silver	Aug 11, 2021, 9:11:13 PM	Region	us-east1 (Sout...	Standard	Aug 11, 2021, 9:11:13 PM
us-east1-compose-001-dda4d9f2-buc...	Aug 12, 2021, 10:00:47 PM	Region	us-east1 (Sout...	Standard	Aug 12, 2021, 10:00:49 PM

Google Cloud Platform   Data Sprints   data studio

Job details   CLONE   DELETE   STOP   REFRESH

Job ID	job-data-sprints-002
Job UUID	5266f58f-2a18-4719-866f-348ec141d73d
Type	Dataproc Job
Status	Succeeded

EDIT

Start time:	Aug 14, 2021, 10:37:32 AM
Elapsed time:	3 min 8 sec
Status:	Succeeded
Region	us-east1
Cluster	cluster-70fe

Output   LINE WRAP: OFF

Google Cloud Platform   Data Sprints   Search products and resources

FEATURES & INFO   SHORTCUT   DISABLE EDITOR TABS   TRIPS_V...

Explorer   + ADD DATA   trips_vendors   QUERY   SHARE

Type to search   SCHEMA   DETAILS   PREVIEW

Viewing pinned projects.

Table schema

data-sprints-322619
  data_sprints
    VW_AVG_TIME_SATURDAY_SUNDAY
    VW_AVG_TRIP
    VW_BIGGEST_VENDORS
    VW_HISTOGRAM_CASH
    VW_SERIES
    trips_vendors

Filter Enter property name or value

Field name	Type	Mode	Policy
dropoff_datetime	STRING	NULLABLE	
dropoff_datetime_year	INTEGER	NULLABLE	
dropoff_datetime_month	INTEGER	NULLABLE	
dropoff_datetime_day	INTEGER	NULLABLE	
dropoff_datetime_hour	INTEGER	NULLABLE	
dropoff_datetime_minute	INTEGER	NULLABLE	
dropoff_latitude	FLOAT	NULLABLE	
dropoff_longitude	FLOAT	NULLABLE	
passenger_count	INTEGER	NULLABLE	
payment_type	STRING	NULLABLE	
pickup_datetime	STRING	NULLABLE	
pickup_datetime_year	INTEGER	NULLABLE	
pickup_datetime_month	INTEGER	NULLABLE	

JOB HISTORY   QUERY HISTORY   Uploads and Data Sp