



Lab Document: tiny.cc/g4g2019-BQGIS-DOC
This Deck: tiny.cc/g4g2019-BQGIS-Deck

BigQuery GIS Training

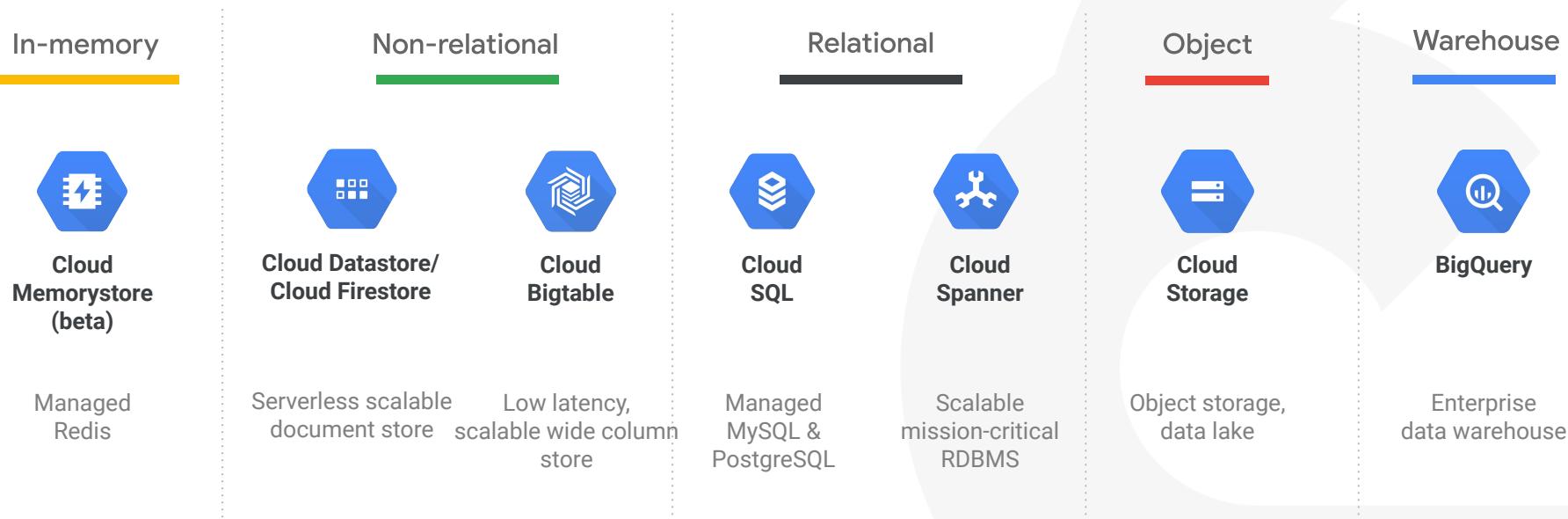
Your SQL and GIS data warehouse in the cloud

2019, Sean Wohltman, Michael A. Entin

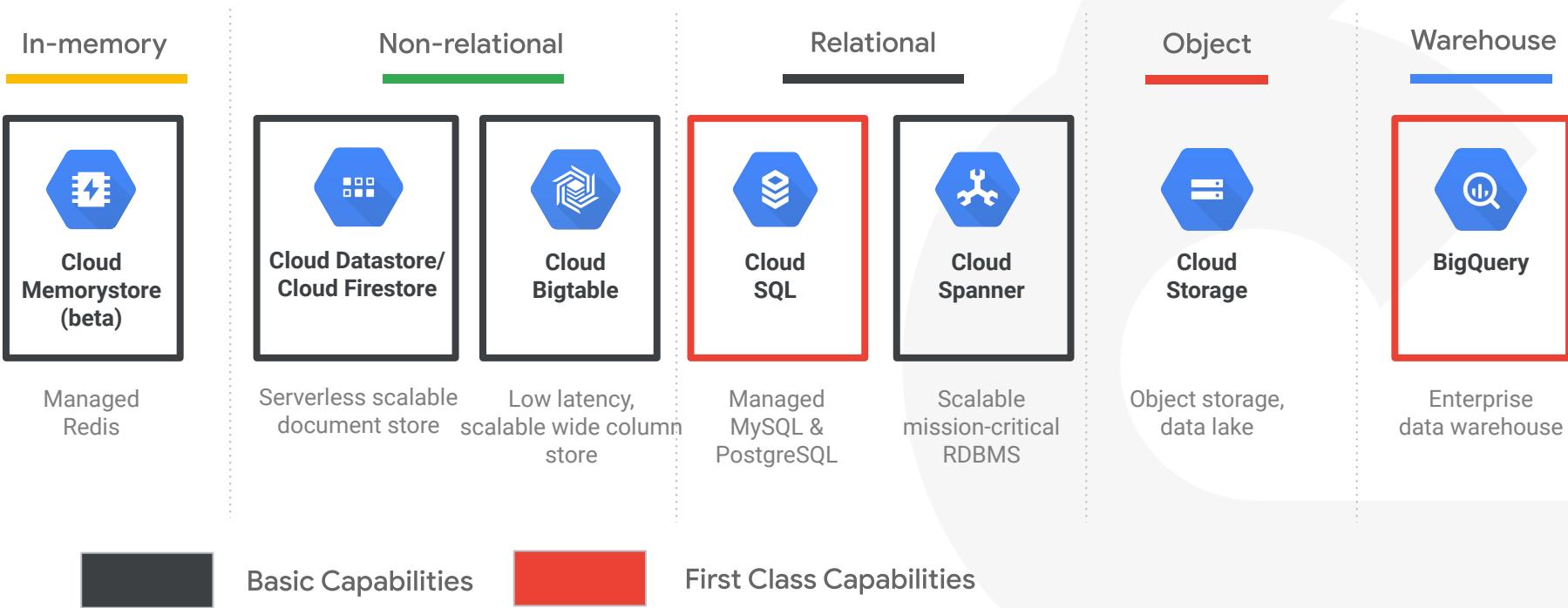
Agenda

- Start Your Engines (Qwiklabs Environment)
- Introducing GCP and BigQuery
- BigQuery GIS
- Demos, hands on learning
- Once You Leave This Class (BigQuery sandbox and Geo Viz)
- Questions?

Portfolio of fully managed data services



Portfolio of fully managed data services



Cloud SQL



Real-time query systems that power interactive applications e.g. on the web and mobile.

BigQuery



Analytical and business intelligence systems driven by queries against large databases.

What is BigQuery?

Google Cloud Platform's enterprise data warehouse for analytics

Convenience of standard SQL

Fully managed and serverless



Petabyte-scale storage and queries

Encrypted, durable and highly available

Real-time analytics on streaming data

```
1 Select language, SUM(views) as views
2 FROM `bigquery-samples.wikipedia_benchmark.Wiki100B`
3 WHERE REGEXP_CONTAINS(title, "G.*o.*o.*g")
4 GROUP BY language
5 order by views desc
6
```

No cached results

Run Save query Save view Schedule query More

This query will process 4.1 TB when run.

Query history

REFRESH

Personal history Project history

Sort by Date Filter queries

BigQuery has had to:

- Read about 1TB of data, then uncompress it to 4TB (assuming ~4:1 compression)
- Execute 100 billion regular expressions with 3 wildcards each
- Distribute 1.25TB of data across the network (1TB compressed for initial read, and 0.25TB for the aggregation)

```
1 Select language, SUM(views) as views
2 FROM `bigquery-samples.wikipedia_benchmark.Wiki100B`
3 WHERE REGEXP_CONTAINS(title, "G.*o.*o.*g")
4 GROUP BY language
5 order by views desc
6
```

No cached results

Run Save query Save view Schedule query More

This query will process 4.1 TB when run.



Query history

REFRESH

Personal history Project history

Sort by Date

To Recreate This You'd Need:

- About 330 100MB/sec dedicated hard-drives to read 1TB of data
- A 330 Gigabit network to shuffle the 1.25 TB of data
- 3,300 cores to uncompress 1TB of data and process 100 billion regular expressions at 1 μ sec per

What is BigQuery GIS?



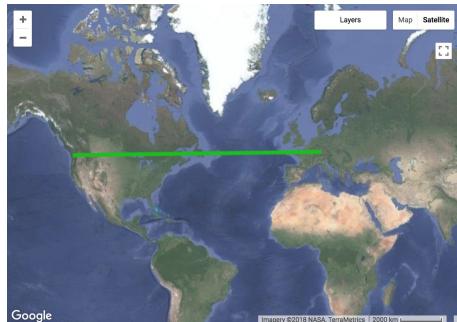
- Vector based spatial analysis
- Geography data type
- Geography [functions](#) - ST_* from SQL/MM standard
- Map [visualization](#)
- Support WKT/WKB, GeoJson field formats

Also in BigQuery GIS

- Public geospatial datasets
 - (US) - counties, zip codes, census, urban areas ...
 - Historic weather from NOAA, hurricanes, lightning strikes ...
- Automatic spatial indices
- Spatial data clustering support
- Spatial JOIN

Geography data type

Geometry



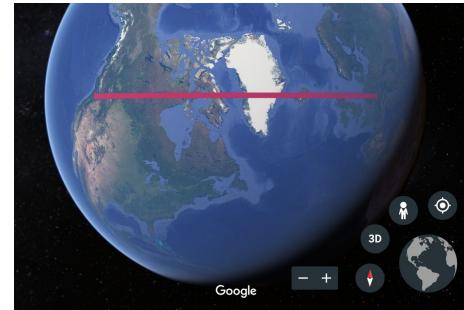
Shortest distance
between 2 points
is a straight line

Geography



Shortest distance is
a great circle.

Geography from space



It is a straight line

WSG84 on a sphere, convert projections outside of BigQuery GIS

Open Geospatial Consortium model

Single
data
type:

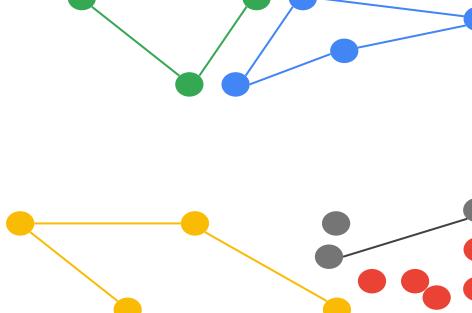
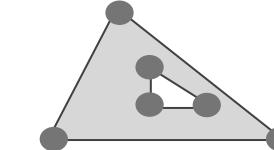
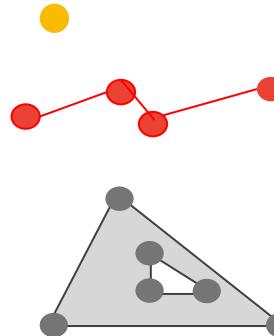
Point

Linestring

Polygon

Multi-polygon

Collections



Can even use it in astronomy



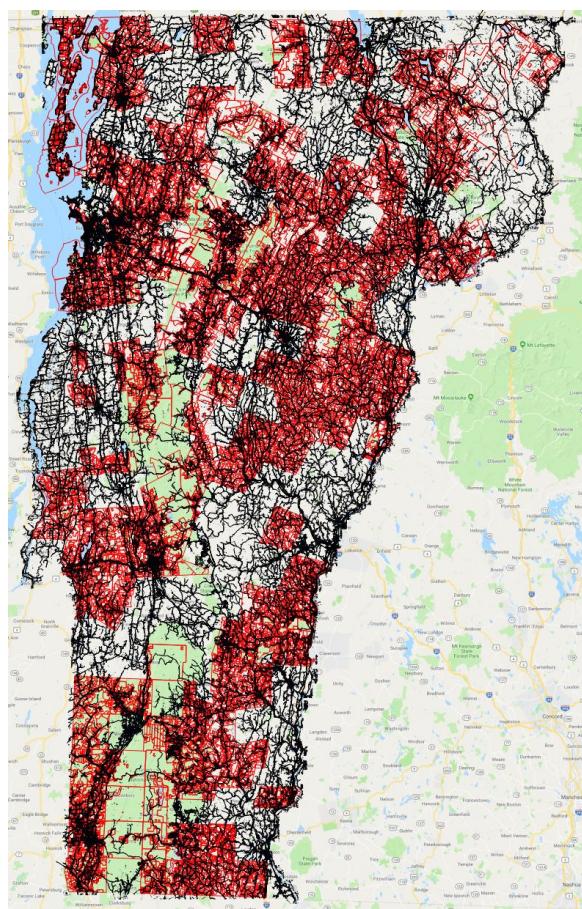
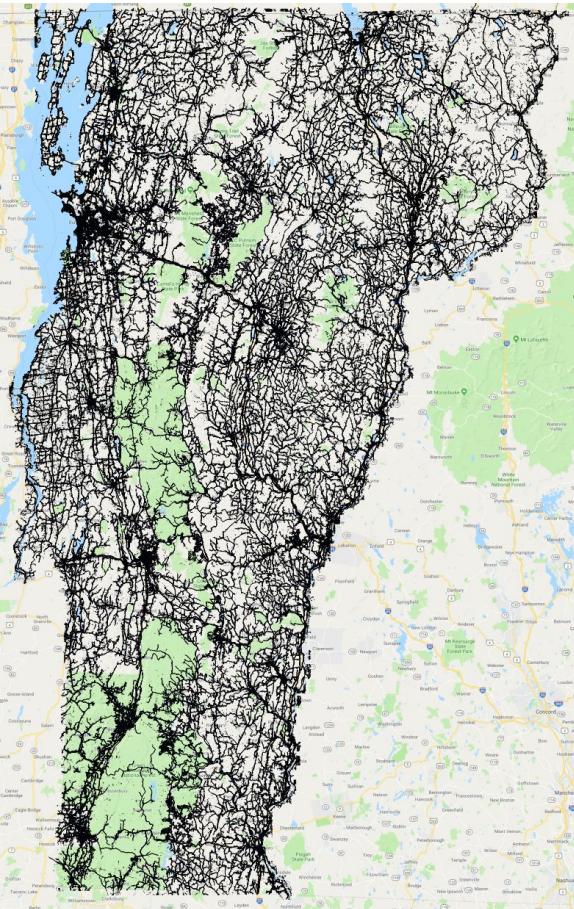
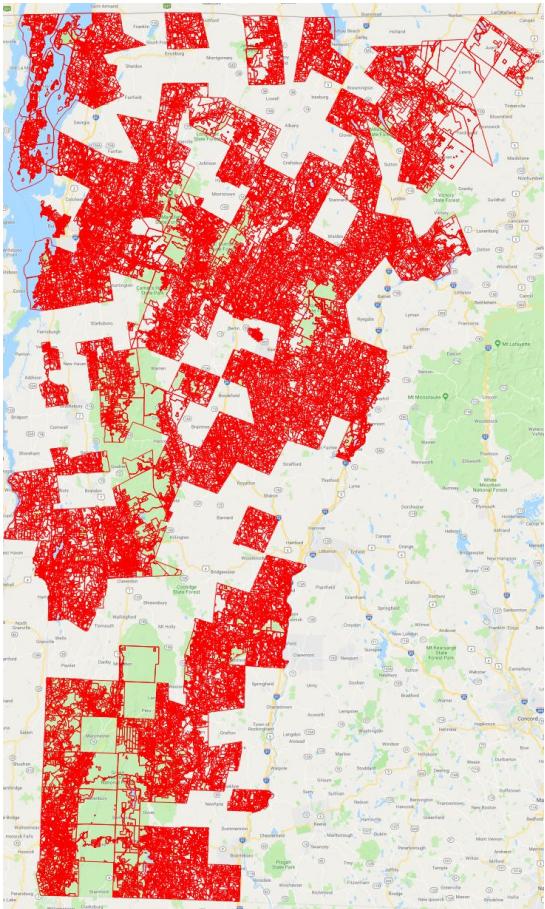
Querying the Stars with BigQuery GIS

<https://cloud.google.com/blog/products/data-analytics/querying-the-stars-with-bigquery-gis>

Ahh..Vermont!



*“There are currently 204,935
Statewide Parcels, how many
of them have impervious
surfaces?”*



#GeoForGood19

Count the number of parcels that are intersected (ST_Intersect) by the impervious surface polygons

```
gisdata=> select count(distinct a.objectid) from "parcels" a, "impervious" b where st_intersects(a.wkb_geometry, b.wkb_geometry);  
count  
-----  
192101  
(1 row)  
Time: 2363688.235 ms (39:23.688)
```

PostGIS on Workstation
12 CPU Core i5
32 GB Ram

#GeoForGood19

*“192,101 of the 204,935 Statewide
Parcels have impervious surfaces.”*

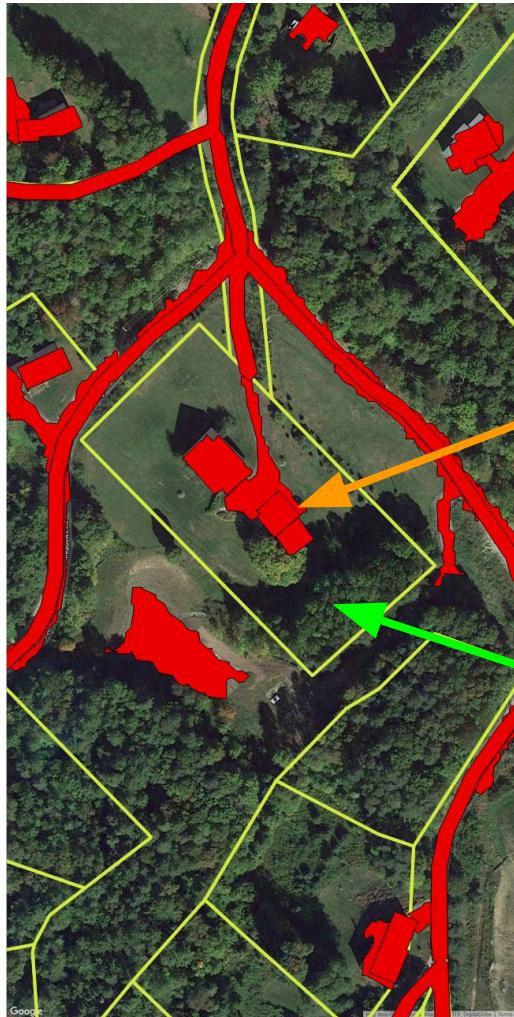


The screenshot shows the Google Cloud Platform BigQuery interface. The left sidebar contains navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, Resources (+ ADD DATA), and a search bar for tables and datasets. The main area is the 'Query editor' with the following query:

```
1 SELECT Count(distinct a.objectid) FROM `vt-solar.vtdata.vtParcels` a , `vt-solar.vtdata.vtImpervious` b where st_intersects(a.the_geom, b.the_geom)
```

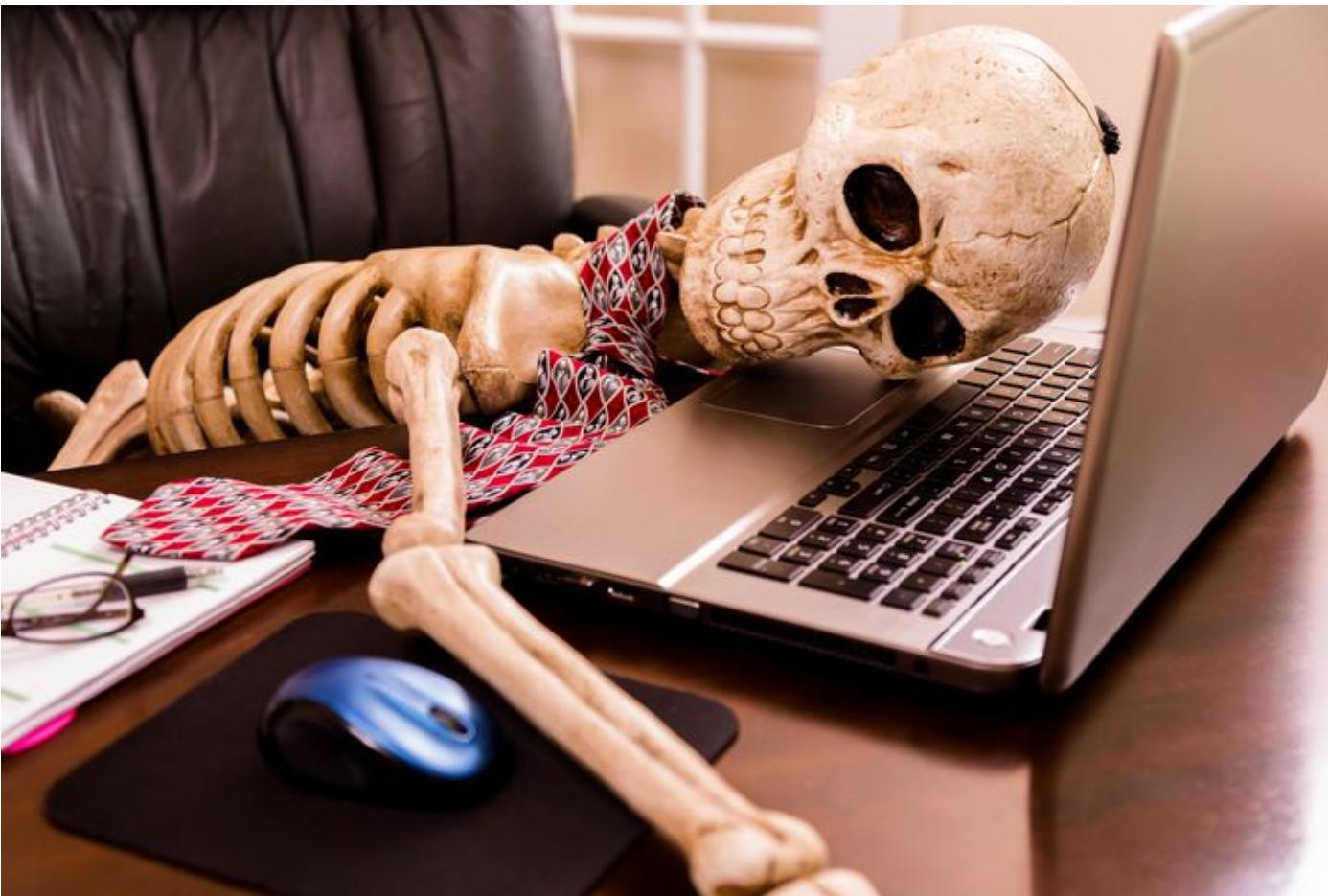
Below the query are buttons for Run, Save query, Save view, Schedule query, and More. A note indicates the query will process 1.4 GB when run. The top right corner shows a user profile icon with a '1' notification.

*“In those 204,935 Statewide
Parcels, what **percent** of
them are impervious
surfaces?”*



#GeoForGood19

*“I didn’t get a screenshot of
attempting this locally, but...”*

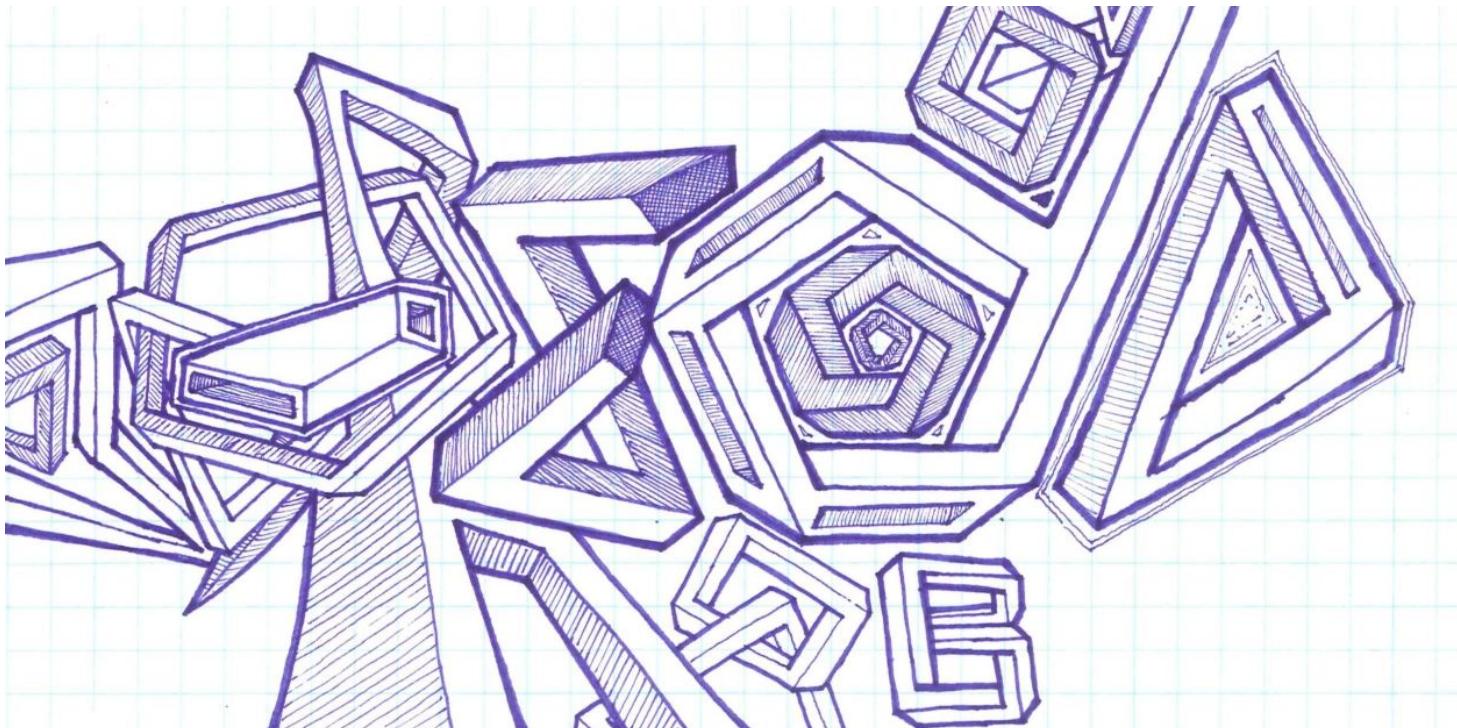


#GeoForGood19



```
1 WITH parcel_impervious_area AS (
2   SELECT p.objectid, Sum(ST_Area(ST_Intersection(im.the_geom, p.the_geom))) AS impervious_area
3   FROM `vt-solar.vtdata.vtImpervious` im
4   JOIN `vt-solar.vtdata.vtParcels` p
5   ON ST_Intersects(im.the_geom, p.the_geom)
6   GROUP BY p.objectid
7 )
8 SELECT p.objectid,p.town,
9 CASE
10   WHEN pia.impervious_area IS NULL
11   THEN 0.0
12   ELSE pia.impervious_area
13 END AS impervious_area,
14 CASE
15   WHEN pia.impervious_area IS NULL
16   THEN 0
17   ELSE round(100 * pia.impervious_area / ST_Area(p.the_geom))
18 END AS impervious_pct,p.the_geom
19 FROM `vt-solar.vtdata.vtParcels` p
20 LEFT JOIN parcel_impervious_area pia
21 ON p.objectid = pia.objectid
22 ORDER BY impervious_pct DESC
```

Handling Bad Geometry



No.	Reason why geometry is non-simple	Description
1	Ring orientation is incorrect (applicable only for polygons)	Polygon which does not self-intersect, but its rings are not oriented correctly.
2	Segment orientation is incorrect	Individual segments are not consistently oriented. The 'to' point of segment "i" should be incident on the 'from' point of segment "i+1".
3	Contains short segments	Some segments are shorter than allowed by the spatial reference system units associated with the geometry.
4	Contains self-intersecting planar parts/rings(for polygons/polylines)	The interior of each part (rings, planar parts) intersects with itself or other parts.
5	Contains unclosed rings (applicable only for polygons)	The last segment in a ring does not have its 'to' point incident on the 'from' point of the first segment.
6	Contains empty parts	Geometry contains empty parts.
7	Contains mismatched attributes	Geometry has mismatched attributes.
8	Contains discontinuous parts	Geometry contains discontinuous parts.
9	Empty Z values (applicable to Z-enabled feature classes)	Geometry is Z-aware but contains NaN Zs.
10	Contains duplicate vertices	Geometry has duplicate vertices.

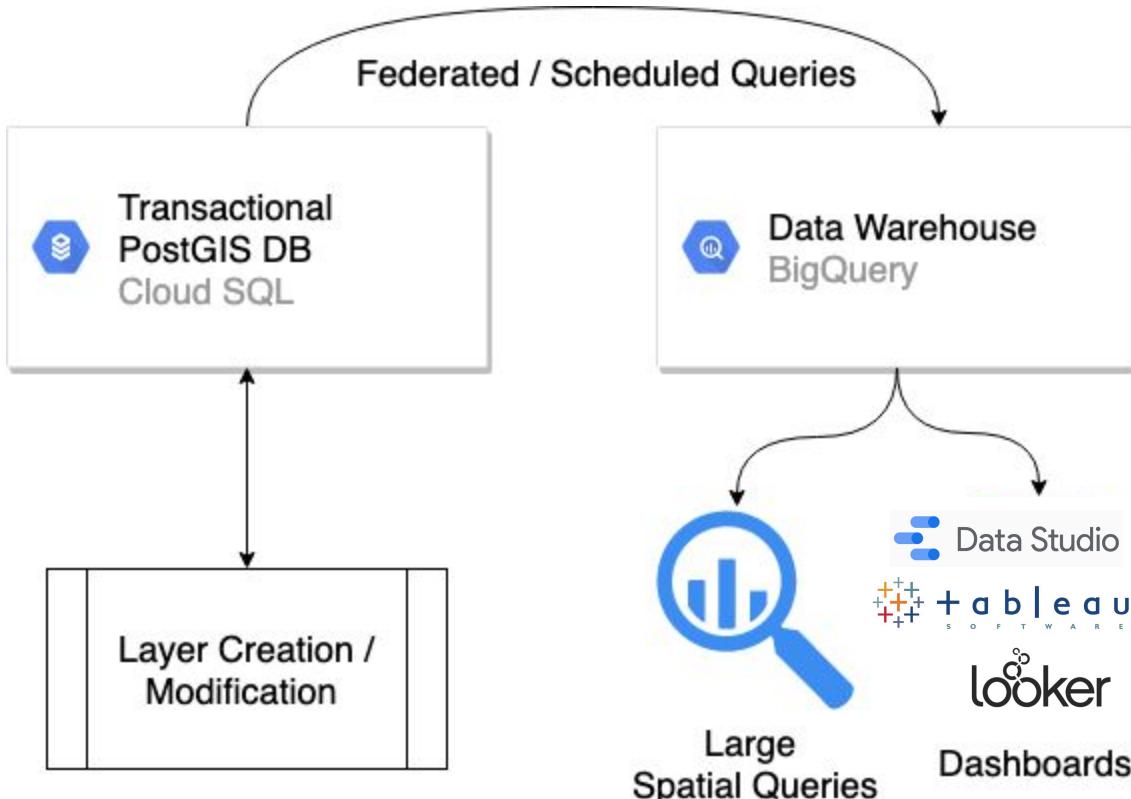


QWIKLABS Section 4

```
1 WITH parcel_impervious_area AS (
2   SELECT p.objectid, Sum(ST_Area(ST_Intersection(im.the_geom, p.the_geom))) AS impervious_area
3   FROM `vt-solar.vtdata.vtImpervious` im
4   JOIN `vt-solar.vtdata.vtParcels` p
5   ON ST_Intersects(im.the_geom, p.the_geom)
6   GROUP BY p.objectid
7 )
8 SELECT p.objectid,p.town,
9 CASE
10   WHEN pia.impervious_area IS NULL
11   THEN 0.0
12   ELSE pia.impervious_area
13 END AS impervious_area,
14 CASE
15   WHEN pia.impervious_area IS NULL
16   THEN 0
17   ELSE round(100 * pia.impervious_area / ST_Area(p.the_geom))
18 END AS impervious_pct,p.the_geom
19 FROM `vt-solar.vtdata.vtParcels` p
20 LEFT JOIN parcel_impervious_area pia
21 ON p.objectid = pia.objectid
22 ORDER BY impervious_pct DESC
```

SAFE

#GeoForGood19



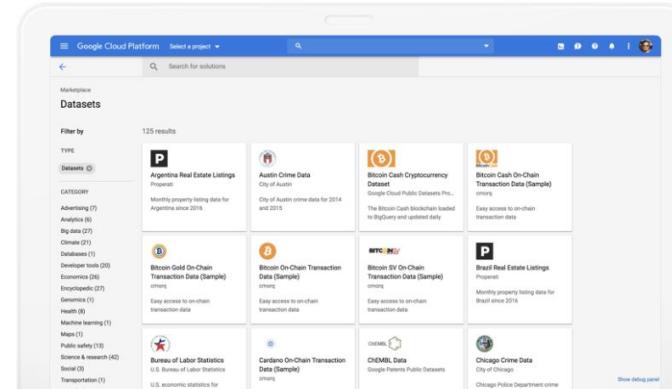
Google Cloud Public Datasets

Uncover new insights with high-demand public datasets.

[View one-pager](#)[Explore public datasets](#)

Explore public datasets in Google Cloud

Google Cloud Public Datasets facilitate access to high-demand public datasets, making it easy for you to access and uncover new insights in the cloud. By analyzing these datasets hosted in BigQuery and Cloud Storage, you can seamlessly experience the full value of Google Cloud with ease.



#GeoForGood19



```
1 WITH parcel_impervious_area AS (
2   SELECT p.objectid, Sum(ST_Area(ST_Intersection(im.the_geom, p.the_geom))) AS impervious_area
3   FROM `vt-solar.vtdata.vtImpervious` im
4   JOIN `vt-solar.vtdata.vtParcels` p
5   ON ST_Intersects(im.the_geom, p.the_geom)
6   GROUP BY p.objectid
7 )
8 SELECT p.objectid,p.town,
9 CASE
10   WHEN pia.impervious_area IS NULL
11   THEN 0.0
12   ELSE pia.impervious_area
13 END AS impervious_area,
14 CASE
15   WHEN pia.impervious_area IS NULL
16   THEN 0
17   ELSE round(100 * pia.impervious_area / ST_Area(p.the_geom))
18 END AS impervious_pct,p.the_geom
19 FROM `vt-solar.vtdata.vtParcels` p
20 LEFT JOIN parcel_impervious_area pia
21 ON p.objectid = pia.objectid
22 ORDER BY impervious_pct DESC
```

After This Class

<https://cloud.google.com/bigquery/docs/gis-intro>

BigQuery sandbox

- Great way to get started with BigQuery
- Free BigQuery tier, no credit card needed
- Up to **10GB user data** stored free
- Up to **1TB of queries a month**
- Full access to public datasets, full performance



Setup BigQuery sandbox:

How-to: <https://cloud.google.com/bigquery/docs/sandbox>



Open browser



Create new or
use existing
account



Click find BigQuery
pin it



Select a project,
NEW PROJECT



<https://console.cloud.google.com/>

SANDBOX Set up billing to upgrade to the full query experience. Learn more **sandbox = Free**

DISMISS UPGRADE

BigQuery FEATURES & INFO SHORTCUTS + COMPOSE NEW QUERY

Query history Saved queries Job history Transfers Scheduled queries BI Engine Resources + ADD DATA

bigquery-public-data

sfdd_service_calls sfpd_incidents street_trees

bikeshare_station_info bikeshare_station_status bikeshare_trips

san_francisco_film_locations san_francisco_sfdd_service_ca... san_francisco_sfpd_incidents san_francisco_transit_muni

Query editor

```
1. SELECT ST_Distance(station_geom, ST_GeogPoint(-122.02, 37.4)) dist, *
2. FROM `bigquery-public-data.san_francisco.bikeshare.bikeshare_station_info`
3. WHERE ST_DWithin(station_geom, ST_GeogPoint(-122.02, 37.4), 1609 * 10)
4. ORDER BY ST_Distance(station_geom, ST_GeogPoint(-122.02, 37.4)) ASC
```

Run Save view Schedule query More

This query will process 89.1 KB when run.

REVIEW RESULTS SAVE RESULTS EXPLORE WITH DATA STUDIO

Query complete (1.3 sec elapsed, 89.1 KB processed)

Job information Results JSON Execution details

results

Row	dist	station_id	name	short_name	lat	lon	region_id	rental_met
1	9637.191176154622	396	Metro Dr at Technology Dr	SJ-C9	37.36767805240811	-121.9187942147255	5	CREDITCA
2	9903.044394732673	401	Sk... Sk... Sk...	SJ-D9	37.36522655191025	-121.9168147444725	5	CREDITCA
3	10367.336226114778	392	Sc... Sc... Sc...	SJ-E9	37.364060462258806	-121.9117319583893	5	CREDITCA
4	10659.475528389497	397	Gi... Gi... Gi...	SJ-E10	37.36186680025918	-121.90931528806688	5	CREDITCA
5	10905.468902224651	395	Kerley Dr at Rosemary St	SJ-F10	37.360854155253826	-121.90683424472809	5	CREDITCA
6	11247.507501803293	393	Asbury St at The Alameda	SJ-I5	37.33800951785354	-121.91942721605302	5	CREDITCA
7	11370.010176248294	391	1st St at Younger Ave	SJ-G10	37.355029582252904	-121.90443634986876	5	CREDITCA
8	11389.789152335701	303	San Pedro St at Hedding St	SJ-H9	37.352601	-121.905733	5	CREDITCA
9	11765.26576523417	288	Mission St at 1st St	SJ-H10	37.3509643	-121.9020161	5	CREDITCA
10	11967.110274896686	402	Raymond Bernal Jr Memorial Park (8th St at Hedding St)	SJ-H11	37.356834352780886	-121.89593575894831	5	CREDITCA
11	12057.787561721017	422	Stockton Ave at Lenzen Ave	SJ-J6	37.33677476275722	-121.90914899110793	5	CREDITCA
12	12124.783165340285	276	Julian St at The Alameda	SJ-K5	37.3322326	-121.9125165	5	CREDITCA

Rows per page: 100 1 - 75 of 75 First page < > >> Last page

BigQuery Geo Viz

App: <https://bigquerygeoviz.appspot.com/>

Help: <https://cloud.google.com/bigquery/docs/gis-visualize>

Use cases:

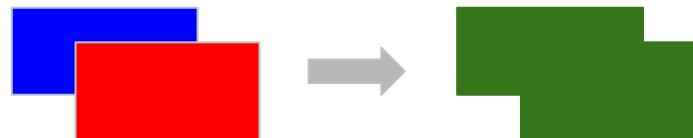
- Ad hoc query visualizations
- Data debugging and cleanup
- Up to 10MB of data



Common stumbling blocks

- 1 **WKT / WKB** functions assume geography on a sphere, geodesic edges, oriented polygons
- 2 Use **GeoJson** when importing planar geometry

- 3 Always dissolved



- 4 Always valid

No ST_MakeValid, fix invalid geometries before loading to BQ

