BigQuery + Google Analytics by Analytics Pros

Overview

Project - Google Developers Console project

Dataset - Group tables and control access to tables. Datasets correspond to GA views.

Table - Tables contain data with the corresponding schema. Google Analytics exports every day's data into a table.

Columns - Features, Attributes - In GA Export Dimensions and Metrics

Rows - Sessions > Hits > Products, Custom Dimensions/Metrics…

Querying

SELECT - Select attributes (dim & met to be returned)

FROM - From which data table(s)

ORDER BY - **<dim|met> DESC|ASC** - order by dimension or metric, descending or ascending

JOIN - Join attributes

AS - Give an alias to an attribute

FLATTEN - Convert repeated fields into an optional field. For example adding CDs to Sessions.

WITHIN - selecting inside nested values

Common Functions

COUNT, SUM, AVG, MAX, MIN, + (addition), - (subtraction), / (division), * (multiplication), % (modulo).
Logical operators: AND, OR, NOT
IF(condition, true_return, false_return)

Example

SELECT totals.visits
FROM [12345678.ga_sessions_YYYYMMDD]
**12345678 = view id*

Query across multiple days

```
FROM TABLE_DATE_RANGE([
  12345678.ga_sessions_],
  TIMESTAMP('2015-05-01'),
  TIMESTAMP('2015-05-15'))
```

Nesting

Sessions ⟶ Hits ⟶ Custom Dimensions

Solution: Flattening, use of within
COUNT(hits.type) WITHIN RECORD AS numberOfHits

Order of commands

SELECT, WITHIN, FROM, FLATTEN, JOIN, WHERE, GROUP BY, HAVING, ORDER BY, LIMIT

Best Practices

1. Only SELECT fields you need
2. Whenever possible use LIMIT
3. Store often queried intermediate tables (save table)
4. Test on smaller dataset
5. Compare results with GA's web ui

BQ Public Samples

publicdata:samples.gsod
publicdata:samples.natality
google.com:analytics-bigquery:LondonCycleHelmet.ga_sessions_20130910

Google Analytics Schema

Session
fullVisitorId
visitNumber
visitStartTime
date
totals.*
trafficSource.*
device.*
customDimensions.*
geoNetwork.*
hits.*

Shell (terminal)

bq shell --project_id=<project id>
show <profile_id>..ga_sessions_20150101
query "SELECT fullVisitorId FROM [<profile_id>.ga_sessions_20150101] LIMIT 10"

Saving

Save Query (requires no space)
Save View (alias for a query)
Save Table (requires space, but offers fast execution and low cost)

Chrome Extension

Streak for bigQuery

Query reference

https://cloud.google.com/bigquery/query-reference

GA + BigQuery cookbook

https://support.google.com/analytics/answer/4419694

GA schema for BigQuery

https://support.google.com/analytics/answer/3437719

AnalyticsPros™

Select all traffic source data

```
SELECT
  trafficSource.*
FROM
  [1234567.ga_sessions_20140415]
LIMIT 1000
```

Count number of sessions (visits)

```
SELECT
  COUNT(fullVisitorId)
FROM
  [1234567.ga_sessions_20140415]
```

Count number of users (visitors)

```
SELECT
  COUNT(DISTINCT fullVisitorId)
FROM
  [1234567.ga_sessions_20140415]
```

Using mathematical functions

```
SELECT
  STDDEV(totals.timeOnSite),
  AVG(totals.timeOnSite),
  SUM(totals.timeOnSite),
  trafficSource.medium
FROM
  [1234567.ga_sessions_20140415]
GROUP BY
  trafficSource.medium
```

Flatten

```
SELECT
  totals.*, hits.hitNumber
FROM FLATTEN([1234567.ga_sessions_20140415], hits)
LIMIT 1000
```

Query across the month of May

```
SELECT
  totals.timeOnSite
FROM
  TABLE_DATE_RANGE([
    12345678.ga_sessions_],
    TIMESTAMP('2015-05-01'),
    TIMESTAMP('2015-05-31'))
LIMIT 1000
```

Subqueries

average number of transactions per user that made a purchase

```
SELECT
  (SUM (total_transactions_per_user) / COUNT(fullVisitorId) )
  AS avg_total_transactions_per_user
FROM
  (SELECT
    fullVisitorId,
    SUM(totals.transactions) AS total_transactions_per_user
  FROM ['Dataset Name' ]
  WHERE totals.transactions IS NOT NULL
  GROUP BY fullVisitorId );
```

Join

```
SELECT *
FROM (SELECT fullVisitorId, visitNumber, customDimensions.value as catId
      FROM FLATTEN([12345678.ga_sessions_20150527], customDimensions)
      WHERE customDimensions.index=2 ) as GA
INNER JOIN [12345678..JoinExample] as JE
ON JE.catId = GA.catId
```

Contact us with questions:

gethelp@analyticspros.com
206-331-4034

5325 Ballard Ave NW, Suite 300 | Seattle, WA 98107

AnalyticsPros™