



Google Analytics and AdSense Data Analysis in BigQuery

Louis Collard, Software Engineer, AdSense Clancy Childs, Product Manager, Google Analytics



Designing Data Access for the "Average" User

Google Analytics	S		Sett	ings My Account	Sign out	Google AdSens	se								0 Alerts Sign out
www.googlestore.com - ht www.googlestore.com	ttp://www.googlestore.com - Reporting Cu	ustomization		Admin	Help 🕑	Home My ads	Allow & b	lock ads Perfor	mance reports						Help
Q, Find reports & more	Browser & OS		Mar 22, 20	13 - Apr 21, 2	2013 -	Common reports Entire account by day		Targeting types	Save report Set as def	ault report				All tin Nov	e 29, 2004 – Apr 22, 2013
MY STUFF	Advanced Segments Customize Email Export -	Add to Dashboard Shortcut				Entire account by week	ı	E-limitedminer	A.d	C	011-1			000	Adverseration
Dashboards	• % of visite: 100.00%					Products		\$348.27	14 466	100.00%	1.8	68 12	91%	\$0.19	\$24.08
Shortcuts	- · ·					Custom channels		φ0+0.2.1		100.0070	1,0			\$ 0110	
Intelligence Events	Explorer	Cool Cot 4 . Ecommono				URL channels									
	Site Usage Goal Set 1 Goal Set 2 Goal Set 3	Goal Set 4 Ecommerce	-			Sites									
STANDARD REPORTS	Visits VS. Select a metric		Day	Week Month	4 ••	Owned sites BETA		Interest-based							
🚺 Real-Time	• Visits					Platforms									
🚢 Audience	10,000					Ad units									
Overview						Ad sizes		Placement							
▶ Demographics	5,000					Ad types									
▶ Geo						Ad networks									
Behavior	Mar 29	Apr 5	Apr 12			Bid types		Contextual							
Browser & OS	Visits Pages / Visit	Avg. Visit Duration	% New Visits	Bounce F	Rate										
Network	215,964 3.89	00:01:24	80.40%	20.47	%	QUICK reports		0		4,000		8,000		12,000	
▶ Mobile	% of Total: 100.00% (215,964) Site Avg: 3.89 (0.00%)	Site Avg: 00:01:24 (0.00%)	Site Avg: 80.40% (0.00%)	Site Avg: 20.47	7% (0.00%)	Saved reports		Targeting types 👻 »	Add dimension 👻						
▶ Custom	Primary Dimension: Browser Operating System Screen Reso	olution Screen Colors Flash	Version Other -					Targeting typ	pe Ad request	s Coverage	Clicks	Ad request CTR	CPC	Ad request RPM	Estimated earnings ↓
Visitors Flow	Plot Rows Secondary dimension V Sort Type: Default	•	Q advanced	⊞ © Ξ '	72 3777			Contextual	13,59	1 100.00%	1,528	11.24%	\$0.22	\$25.05	\$340.49
Traffic Sources	Browser	Visits ↓ Pages Visit	/ Avg. Visit Duration	% New Visits	Bounce Rate			Interest-base	d 9	5 100.00%	1	1.05%	\$0.00	\$9.97	\$0.00
Content	1. Chrome	108,751	4.45 00:01:20	87.96%	18.06%			Averages	4,82	2 —	622	-	-	-	\$116.09
	2. Internet Explorer	28.004	3.44 00:02:23	91.95%	24.18%			Totals	14,46	6 100.00%	1,868	12.91%	\$0.19	\$24.08	\$348.27
Conversions								Search Targeting ty	pes Go			Got	o page:	1 Show rows: 50 -	1 - 3 of 3 14 4 14 14

Designing Data Access for Data Scientists



BigQuery: Google's Big Data Analysis Engine



2013-05-01 3327.0 Tablets 1132581 1904126 10241.0 2 2013-05-01 High-end mobile devices 3 2013-05-02 Desktop 7741152 24573.2 2013-05-02 Tablets 7474616 32643.8 4 2013-05-02 High-end mobile devices 5354473 13257.0 5

Huge Scale

Interactive Processing

Cost Effective

Third-Party Ecosystem

👬 + a b | e a u



QlikView



METRIC INSIGHTS



AdSense Data in BigQuery

AdSense Data in BigQuery

Tables

DailyReport

Daily AdUnit Report

Daily**CustomChannel**Report

Daily**UrlChannel**Report_

Daily**Domain**Report

Common Dimensions

Date Ad Client Product Ad Format Bid Type Targeting Type Platform Type Country **Metrics** Ad Requests Matched Ad Requests Individual Ad Impressions Clicks Earnings

Compared to the AdSense UI

Increased Flexibility

Request any combination of dimensions and metrics

Scalable

Query hundreds of thousands of rows

Live data

Matches the AdSense UI and API, updates throughout the day

Free to query

No charges for querying the AdSense BigQuery tables

Top Channels

BigQuery

SELECT

```
custom_channel_id,
```

```
SUM(earnings) AS earnings
```

FROM

[google.com:adsense-reports:Reports.DailyCustomChannelReport]
GROUP BY

```
custom_channel_id
```

```
ORDER BY
```

```
earnings DESC
```

```
LIMIT
```

5

Top Growing Channels

BigQuery

```
SELECT
```

```
jan_report.custom_channel_name,
  feb_report.clicks_sum - jan_report.clicks_sum AS clicks_delta
FROM
  (SELECT custom_channel_name, SUM(clicks) AS clicks_sum
   FROM [...DailyCustomChannelReport]
   WHERE date >= '20130101' AND date < '20130201'
   GROUP BY custom_channel_name) AS jan_report
INNER JOIN
  (...WHERE date >= '20130201' AND date < '20130301'...) AS feb_report
ON jan_report.custom_channel_name = feb_report.custom_channel_name
```

ORDER BY clicks_delta DESC LIMIT 5

Third party integration: Tableau

← → 🖪 🛢 🖬 • 🚇	<mark></mark>		> - Abc	豆 🖬 -	Normal 🚽 H	⊨ ∠-		<u>illi</u>	Show Me
Data ÷	Pages	Colur	mns	DAY(date)					
odSense Daily Report				,					
Dimensions 🔠 🔎 👻		Rows		SUM(earnings	;)				
Abc ad_client_id	Filters	Γ							
Abc ad_format_code	date						A I		
Abc ad_format_name								∧ ∧	
Abc bid_type_code				A				$\nabla \Lambda$	$\wedge \wedge$
m country code	Marks		1	$\langle \Lambda \rangle$	$\langle \Lambda \rangle$			∧ \/ \/	
country_code	√ Line ▼			$\Lambda h \Lambda I$				1 V V	
Abc custom channel code				$\nabla V \nabla V$				\ <u>^</u>	٨
Abc custom channel id	🛛 😓 🖉 🕺 Abc 123			V	\times / \times	M D N			
Abc custom_channel_name	Color Size Label	8		V V				V V	$^{\prime}$ $^{\prime}$
😰 date		in i			V	• •			V
Abc platform type code	Detail Tooltip Path	e					γ		
Measures									
#_	😞 custom_channel_na						$-1 \times 1 \times$	7	
# ad_requests			へ	\sim	-			\sim	\sim . I
# clicks			A-						
=# earnings			- 30	\sim					
# individual_ad_impressions			- 1						
# Inatched_ad_requests			1			No to			202 mulle
I onaitude (generated)									292 nuns
=# Number of Records				Nov 12, 12	Nov 27, 12	Dec 12, 12	Dec 27, 12	Jan 11, 13	Jan 26, 13
# Measure Values						Day of da	ite		

Getting Started

Beta program open to all AdSense publishers from today!

Set up a BigQuery project for your AdSense user

http://bigquery.cloud.google.com

Start querying data

http://developers.google.com/adsense/bigquery

For support or to report problems

http://developers.google.com/adsense/community



Google Analytics and BigQuery

Google Analytics Data in BigQuery

What are we announcing today?

- Google Analytics data will be made available in BigQuery for Google Analytics Premium clients.
- Targeting a full launch by September 2013.
- Opening up a group for developers that are interested in working with this data when it is available.

Google Analytics Data in BigQuery

- A daily export from a Google Analytics profile into a BigQuery dataset.
- Each row represents a session/visit in the Google Analytics profile.
 Session-level data traffic source, custom dimensions, etc.
 Hit-level data sequenced hits included within it's session/row
- Dimensions and metrics map to the names used in the Google Analytics Core Reporting API.
- At launch, this will contain first-party (site) data.
 - Some post-processed data like Geography or AdWords may not be available at launch.

Google Analytics Data in BigQuery - Developer's Group

For information about Google Analytics Premium:

http://google.com/analytics/premium

For future updates, please complete the form:

http://goo.gl/QJR9Y







Duncan McKie

Senior Business Analyst eBay Classifieds Group



eBay Classifieds Group - Our Business





> MARKTPLAATS.





eBay Classifieds Group - Our Data

- Petabytes of internal data:
 - site databases
 - CRM
 - data warehouses
 - marketing systems

- External data:
 - Google Analytics
 - \circ AdSense
 - payment providers
 - advertising data
 (DoubleClick,
 AdWords, other
 networks)

eBay Classifieds Group - Our Challenges

- Integrating data from different providers and making it available to large numbers of analysts across the business.
- Evaluating large volumes of data to investigate complex issues.
- Providing an accessible source of insight to enable non-power users to make data-driven decisions.

404 Page not found \longrightarrow Sad face!

Sorry, the page you're trying to load doesn't seem to exist.

Back to Gumtree.com

Debugging Errors

- Errors such as 404 & 500 lose users!
- Difficult to find the cause of errors using Google Analytics UI
- Hit-level data in BigQuery allows us to examine user journeys

Debugging Errors: Step 1

```
BigQuery
```

SELECT

```
visitorId AS visitorId1
```

```
,visitId AS visitId1
```

```
,hits.customVariables.customVarValue AS errorType
```

```
,hits.hitNumber
```

FROM

```
[gumtree-uk-ga:google_analytics.sessions_20130421] AS t1 WHERE
```

```
hits.customVariables.index = 1
```

AND REGEXP_MATCH(hits.customVariables.CustomVarValue,'404|500');



Debugging Errors: Step 2

```
SELECT
                                                                                  BigQuery
  errorType.visitorId1
  ,errorType.visitId1
  ,errorType.errorType
  ,MAX(previousHit.hits.hitNumber) AS hitNumber
FROM
  (SELECT
    . . .
    ) AS errorType
INNER JOIN EACH
    (FLATTEN ([gumtree-uk-ga:google_analytics.sessions_20130421], hits.hitNumber))
        AS previousHit
    ON errorType.visitorId1 = previousHit.visitorId
    AND errorType.visitId1 = previousHit.visitId
WHERE
  errorType.hits.hitNumber < previousHit.hits.hitNumber</pre>
GROUP EACH BY 1,2,3;
```

Debugging Errors: Step 3

SELECT

pageName.hits.hour AS hour

,hitFinder.errorType AS errorType

,pageName.hits.customVariables.customVarValue AS pageName

,COUNT(*) AS errorCount

FROM

(FLATTEN([gumtree-uk-ga:google_analytics.sessions_20130421],hits.hitNumber)) AS pageName INNER JOIN EACH

(SELECT

```
) AS hitFinder
ON pageName.visitorId = hitFinder.visitorId1
AND pageName.visitId = hitFinder.visitId1
AND pageName.hits.hitNumber = hitFinder.hitNumber
WHERE
pageName.hits.customVariables.index = 1
GROUP EACH BY 1,2,3
ORDER BY 4 DESC;
```

BigQuery

Debugging Errors: The Results

Query Results

Row	hour	errorType	pageName	errorCount
1	21	Error-404	Search	1775
2	21	Error-404	VIP	1469
3	20	Error-404	Search	1458
4	22	Error-404	Search	1416
5	21	Error-404	Listing	1383



- Can investigate any sequence of hits
- No need to set up funnels in advance

Segmenting Onsite Activity



- Each post has attributes (category, location, etc.) used across business to segment user activity
- Segments can be complex and dynamic, changing with business needs



• Import lookup tables into BigQuery as CSV or JSON

Row	location_id	location_level	location_level_1	location_level_2	location_level_3	location_level_4	location_name
1	881	4	Midlands	West Midlands	Birmingham	Hodge Hill	Hodge Hill
2	870	4	Midlands	West Midlands	Birmingham	Tile Cross	Tile Cross
3	546	3	South	East Sussex	Brighton	null	Brighton
4	905	4	Midlands	West Midlands	Birmingham	West Bromwich	West Bromwich
5	836	4	Midlands	Nottinghamshire	Nottingham	Dunkirk	Dunkirk
6	939	4	Midlands	West Midlands	Birmingham	Winson Green	Winson Green

SELECT

hits.customVariables.customVarValue AS location_level_2
,visitId
,visitorId
,COUNT(hits.hitNumber) AS pageviews
,SUM(totals.bounces) AS bounces
FROM
 [gumtree-uk-ga:google_analytics.sessions]
WHERE

BigQuery

```
hits.customVariables.index = 4
GROUP EACH BY 1,2,3;
```

SELECT

```
location_level_2
```

```
,location_level_1 AS tv_region
```

FROM

```
[gumtree-uk-ga:gumtree_uk_metadata.location]
```

WHERE

```
location_level_1 IN ('North West',
 'South',
 'Midlands',
 'South West')
GROUP BY 1,2
```

BigQuery

```
BigQuery
SELECT
    location.tv_region
    ,COUNT(DISTINCT CONCAT(STRING(ga.visitorId), '-',STRING(ga.visitId))) AS visits
    ,COUNT(DISTINCT ga.visitorId) AS visitors
    ,SUM(ga.pageviews) AS pageviews
    ,SUM(ga.bounces) AS bounces
FROM
    (SELECT
        . . .
     ) AS ga
INNER JOIN
    (SELECT
        . . .
     ) AS location
    ON ga.location_level_2 = location.location_level_2
GROUP BY 1;
```

Segmenting by TV Region: The Results

Query Results

Row	location_location_level_1	visits	visitors	pageviews	bounces
1	North West	137010	99368	672986	25279
2	South	110954	85544	523778	20578
3	Midlands	217683	161621	1105173	38502
4	South West	93297	70974	445084	16032

- No need to individually query each segment
- No pre-aggregation or sampling of data

A|B Testing - AdSense & Analytics

- A|B testing allows us to improve our implementation
- Interested in impact on whole user experience
- Optimisation metrics (e.g. Clicks/Visit) derived from various sources



A|B Testing - AdSense & Analytics: Step 1

SELECT

```
custom_channel_name AS adSenseGroup
,SUM(ad_requests) AS ad_requests
,SUM(matched_ad_requests) AS matched_ad_requests
,SUM(clicks) AS clicks
,SUM(earnings) AS earnings
FROM
      [874385582184:Reports.DailyCustomChannelReport]
WHERE
```

```
REGEXP_MATCH(ad_client_id,'gumtree-com')
```

GROUP BY 1

BigQuery

A|B Testing - AdSense & Analytics: Step 2

BigQuery

SELECT

hits.customVariables.customVarValue AS adSenseGroup

,COUNT(DISTINCT CONCAT(STRING(visitorId), '-', STRING(visitId))) AS visits

,COUNT(DISTINCT visitorId) AS visitors

,COUNT(hits.hitNumber) AS pageviews

FROM

[gumtree-uk-ga:google_analytics.sessions]

WHERE

```
hits.customVariables.index = 42
GROUP BY 1
```

A|B Testing - AdSense & Analytics: Step 3

BigQuery

SELECT

ga.adSenseGroup

,ga.visits

,adSense.matched_ad_requests

,FLOAT(adSense.clicks/adSense.matched_ad_requests) AS CTR

,FLOAT(adSense.matched_ad_requests/ga.visitors) AS matched_ad_requests_per_visitor

```
,FLOAT(ga.pageviews/ga.visits) AS pageviews_per_visit
```

FROM

(SELECT

• • •

) AS ga

INNER JOIN

(SELECT

• • •

) AS adSense

ON ga.adSenseGroup = adSense.adSenseGroup

A|B Testing - AdSense & Analytics: The Results

Query Results

Row	test_group	CTR	matched_ad_requests_per_visitor	pageviews_per_visit
1	testGroup1	0.020860557309110735	7.107795596644189	9.720080938168486
2	testGroup2	0.022982172537055166	7.505981406371053	10.323799771939214
3	testGroup3	0.02324803478842616	7.078607463976023	10.050262812089356

- Able to evaluate all factors in one place
- Using same Custom Dimension means one setup for all tests

In Summary / Next Steps

Get up and running with BigQuery: http://bigquery.cloud.google.com

Google Analytics in BigQuery

- Register for more information: http://goo.gl/QJR9Y
- Targeting September Launch for Google Analytics Premium users

AdSense in BigQuery

- Beta available now!
- Step-by-step instructions on getting started: http://developers.google.com/adsense/bigquery



Thank You and Questions?



BigQuery: http://bigquery.cloud.google.com AdSense: http://developers.google.com/adsense/bigquery Google Analytics Premium Data: http://goo.gl/QJR9Y

