Got it, this is helpful, thank you. Quick question, is a place I can look up for campaign name?

Also, what tables I can look for Bids and ASIN info at campaign level.

Thanks
Sharath
From: Asham, Mina <minaasha@amazon.com>
Sent: Tuesday, February 19, 2019 12:42 AM
To: B S, Sharath <sharab@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Mokarim, Zohair <zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Sharath,

We stopped hashing the spadid, hence it's only available from the 13th, it's joinable with the sp_ads.ad_id field, here is a sample query:

```
SELECT campaign_id
FROM sp_ads
JOIN attributedeventsdailymanifest ON
attributedeventsdailymanifest.spadid = sp_ads.ad_id
WHERE sp_ads.day = '2019-02-14'
AND   attributedeventsdailymanifest.day = '2019-02-14'
LIMIT 10;
```
Cheers,
Mina

From: B S, Sharath <sharab@amazon.com>
Sent: 19 February 2019 02:46
To: Asham, Mina <minaasha@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Mokarim, Zohair <zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Mina,

Thank you for emailing, the data in **spadid** column are hash function, what column I should use to join with **sp_ads?**

Thanks
Sharath
From: Asham, Mina <minaasha@amazon.com>
Sent: Monday, February 18, 2019 1:15 AM
To: McBryan, Tony <mcbryant@amazon.co.uk>; Mokarim, Zohair <zohaim@amazon.com>; B S, Sharath <sharab@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>
Subject: RE: Retargeting metrics deck

Hey Sharath,

Just adding that we should now have campaign-level data, feel free to start using the **spadid** column in **attributedeventsdailymanifest** to join with **sp_ads** table and extract campaign ids. This is only available from the 13th of Februrary though (the field did exist in earlier days but it does not have the proper value).

Cheers,
Mina

From: McBryan, Tony <mcbryant@amazon.co.uk>
Sent: 13 February 2019 11:38
To: Mokarim, Zohair <zohaim@amazon.com>; B S, Sharath <sharab@amazon.com>; Asham, Mina <minaasha@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>
Subject: RE: Retargeting metrics deck

Hiya,

1) **Data pipelines**

We'll look to getting this work scheduled. We might wait another week or two to make sure the thrash on the requirements of the WBR settles down first though (i.e. active advertisers definition etc).

2) **Week 6**

2/9 looks like it's available in the data now. I think this got stuck behind https://t.corp.amazon.com/AIMOncall-252 although it should be noted it was never out of our SLA. (i.e.

it's not unusual for the data to take this long). Moving forward we've pushed a faster version of the datafeed but the SDP data is currently only of WBR quality for days after 2/10 in that table so isn't quite ready to be used.

3) Active advertisers

Use this. I'm not entirely sure this is where the definition of active advertisers will land, there are probably some more refinements we could do. This isn't exactly the same as the query I used last time as this has some extra refinements but it's not possible to trivially backfill this as spbetaadvertisers isn't versioned and has had a lot of advertisers added to it.

```
WITH
  da_campaigns AS (
    SELECT *,
    (COALESCE (end_date, '') = '' OR end_date > '2019-02-02 23:59:59Z') and (start_date <= '2019-02-02 23:59:59Z') as withinDate
    FROM da_athena.sp_campaign WHERE (venues = '2' OR (bitwise_and(CAST(venues AS BIGINT), 2) > 0 AND advertiser_id IN (SELECT advertiser FROM da_athena.spbetaadvertisers)))
  )
SELECT count(DISTINCT advertiser_id) as count FROM da_campaigns WHERE day = '2019-02-02' and status = '0' and marketplace_id = '1' and withinDate
```

Tony

From: Mokarim, Zohair <zohaim@amazon.com>
Sent: 13 February 2019 06:56
To: B S, Sharath <sharab@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Asham, Mina <minaasha@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Tony/Mina

We tried re-running the metrics report for Week 6 and still see no data for Saturday 2/9. Understand from the below thread that we have a lag in the data input – how can we reduce this to address this problem. Given the WBR timings, we will only have 6 days worth of data to present tomorrow. We had previously utilized ManifestV2 as it had complete data but for this week even that is incomplete. How can we address this going forward ? As always, really appreciate your help with all this!

Regards
Zohair

From: Mokarim, Zohair
Sent: Tuesday, February 12, 2019 4:31 PM
To: B S, Sharath <sharab@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Asham, Mina <minaasha@amazon.com>

Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Tony/Mina

Bringing this thread back up. In addition to the below data pipelines, can you please also advise on how we are arriving at the active advertisers number of 2292 (data source and query) so that we can start reporting on this week in week out. Thanks once again for all your help.

Regards
Zohair

From: B S, Sharath <sharab@amazon.com>
Sent: Thursday, February 07, 2019 11:37 PM
To: McBryan, Tony <mcbryant@amazon.co.uk>; Asham, Mina <minaasha@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Tony, Mina,

Thank you for the response it is helpful,

We were successfully able to report the metrics in the WBR this week, here is the joined query I used to report/dashboard: https://paste.amazon.com/show/sharab/1549610850

To automate the extraction of data, if you can set up a pipeline to drop the output of the query either via EDX or Andes or S3 (whichever works for you) and grant the right permission, I will be able to load into our Redshift cluster automate at our end. We will get back to the campaign level data as metric needs evolves.

Thank you again for all the support.

Just in case needed Our AWS account and role: arn: aws: iam:: **006877342108**: role/Redshiftunloadcopy
Thanks
Sharath
From: McBryan, Tony <mcbryant@amazon.co.uk>
Sent: Tuesday, February 5, 2019 4:04 AM
To: Asham, Mina <minaasha@amazon.com>; B S, Sharath <sharab@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair

<zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

- @Tony: As attributedEventsDailyManifest has 4 days lag, and we create our reports on Tuesday PST (for now), which will delay to report last week's data, hence I was asking for the new table you mentioned which is faster.

Tony can confirm, but it is a work in progress at the moment, people are used to the lag in reports though so I don't think it's a problem.

Yup, we're sad too. When we fix up the schema and vend it to you we can keep you on the fastest data sources. The SLA will always be at least 2-3 days as that is the fastest cadence we get the input data at.

- What is the best source to obtain active advertiser impressions data for the day/week?
  If you just want impressions without conversions, then you can use this table: advertisingeventsdailymanifest

Depends on what you mean by active. Grouping your earlier query on attributedeventsdailymanifest / m3conversions by advertiserid will give you advertisers with impressions/clicks/conversions.

- What is the difference between the ASINs vs Ads fields (ASIN is the eligible in the advertisers' catalog while ads is the number of ads that we show?)

Will leave this one to Tony.

These are all campaign config metrics using the SP Axiom concepts. An ad is an ASIN in an ad group (i.e. an ASIN can be in multiple ad groups so has multiple "ads").

- To get the total campaign budget should we multiply the daily budget by the campaign run time? And how do you calculate such budget for evergreen campaigns?

Roughly, but that wouldn't account for the budget changing during that runtime. We don't have that audit information in Athena, the closest you can do is getting the budget for every day from the sp_campaign

I think this isn't entirely trivial to reason about because there isn't a real total budget on evergreen. You can (somewhat) reason about budgets within a time period (i.e. weekly) which would be a weighted combination of the budgets live time (i.e. $200/day budget for 33% of the time and $400/day for 66% of the time = (200*033+400*.66)*7 = $2333/wk.

Tony

From: Asham, Mina <minaasha@amazon.com>
Sent: 01 February 2019 10:00
To: B S, Sharath <sharab@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair

<zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

Comments inline.

-Mina

From: B S, Sharath <sharab@amazon.com>
Sent: 01 February 2019 06:20
To: Asham, Mina <minaasha@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

Got it , thank you for that

We were trying to take a deeper look at the data, we have few more detailed level questions as below.

1. @Mina: Regarding venue you mentioned "So "venues & 2 = 2" reads as "extract second bit from the bit array, and make sure it's set to 1. The queries you shared had venues = 2, so ideally for offsite data I should change it to 1 right?(reconfirming it)

Traditionally we ran hybrid campaigns (SP Onsite + Remarketing Offsite) and we still have some of them active, so you options are "venues = 2" (offsite only) or "venues & 2 = 2". If you change the number to 1 that means you are extracting bit-0 which can only mean onsite (1 = binary 100, 2 = binary 010, etc…).

2. @Tony: As attributedEventsDailyManifest has 4 days lag, and we create our reports on Tuesday PST (for now), which will delay to report last week's data, hence I was asking for the new table you mentioned which is faster.

Tony can confirm, but it is a work in progress at the moment, people are used to the lag in reports though so I don't think it's a problem.

3. Currently based on status code 0,1,2 we are identifying active , paused, archived in the query. However, can you please provide us with the logic/definition used to capture Active, Paused, Archived status in this table ?

Not sure I understand the question, which table are we talking about? In general the "status" is the status of the campaign in Axiom's DB. Now there is "status" in the data store and there is "calculated status" that appears in the UI, the latter would include out of budget and out of date range, etc…

4. What is the best source to obtain active advertiser impressions data for the day/week?

If you just want impressions without conversions, then you can use this table: advertisingeventsdailymanifest

5. Going through the Active Campaign Status we noticed that out of 45 campaigns that have an end date specified, 25 of those dates are in the past but the camping status is still set on Active (example is below), can you please help us understanding why it is the case?

a.

advertiser_id	campaign_id	status	creation_date	start_date	end_date
28467010002	2.00001E+14	active	7/30/2018 15:59	8/1/2018 4:00	8/17/2018 6:59

Yes, that is expected, the status reflects the advertiser manually paused or archiving the campaign. There are a lot of campaigns with end data in the past where the status is 0, the calculated status will be paused/ended though.

6. In the Paused campaigns, we noticed that some of the End Dates are in the past. If the campaign end date is already past due, shouldn't the status of the campaign be set on Archived/Ended?

Same as above.

7. To report on campaigns created for a particular week, we'll be utilizing the creation date field (date falling in the reporting week), is there anything else that we need to consider?

That makes sense, that is what I have used before. Only thing to note is that the date is in UTC.

8. What is the difference between the ASINs vs Ads fields (ASIN is the eligible in the advertisers' catalog while ads is the number of ads that we show?)

Will leave this one to Tony.

9. To get the total campaign budget should we multiply the daily budget by the campaign run time? And how do you calculate such budget for evergreen campaigns?

Roughly, but that wouldn't account for the budget changing during that runtime. We don't have that audit information in Athena, the closest you can do is getting the budget for every day from the sp_campaign

Thanks
Sharath

From: Asham, Mina <minaasha@amazon.com>
Sent: Thursday, January 31, 2019 4:49 AM
To: McBryan, Tony <mcbryant@amazon.co.uk>; B S, Sharath <sharab@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>
Subject: RE: Retargeting metrics deck

> The sp_* tables I think are a dump of Axiom but I'd leave Mina to confirm.

That is right, Axiom does a daily dump using its Vulcan client to EDX, and we sync that to our Athena cluster:
https://w.amazon.com/index.php/ProductAds/Development/CampaignPipelinePlatform/EDX

-Mina

From: McBryan, Tony <mcbryant@amazon.co.uk>
Sent: 31 January 2019 11:10
To: B S, Sharath <sharab@amazon.com>; Asham, Mina <minaasha@amazon.com>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Walton, Chris <walton@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>
Subject: Re: Retargeting metrics deck

We won't remove the attributedEventsDailyManifest version 1 table at all (at least not for a long long time). There are some other things joined in there (irrelevant for SDP) which we need to keep around.

The M3 conversions tables are sourced from MADS attribution pipeline (https://w.amazon.com/index.php/Advertising/D16G/Platform/Analytics/Projects/M3).

The attributed events tables are sourced from a join of the bids we make, the impressions we win, the clicks we drive (and their Ad Revenue) and consumer marketing value (OPS).

The sp_* tables I think are a dump of Axiom but I'd leave Mina to confirm.

Tony

From: B S, Sharath
Sent: 31 January 2019 07:17
To: McBryan, Tony; Asham, Mina
Cc: Blums, Martins; Certini, Luca; Birkett, Andrew; Balgi, Sameer; Hope, Emily; Walton, Chris; Mokarim, Zohair
Subject: RE: Retargeting metrics deck

Wow, thank you Mina and Tony for such a detailed information.

@Tony: I completely understand the issue with change management, I am currently flexible to any such changes since we are in learning phase. 'If possible, ' please notify me if any current tables provided changes, before we automate. Also can you please share the name of replacement table to attributedEventsDailyManifestV2, just in case attributedEventsDailyManifestV1 is not populated before next week WBR.

@Mina/ Tony: One follow up question since I am curious now - What is the data source of these tables? I know Mina mentioned 'sp_* tables are all daily snapshots of SSPA DB' is SSPA DB is same as ADW?

@Mina: Sure, next time if I don't see the data, first I will try by clicking the three dots next to the table and then "Load partitions".

Thanks
Sharath
From: McBryan, Tony <mcbryant@amazon.co.uk>
Sent: Wednesday, January 30, 2019 3:54 AM
To: Asham, Mina <minaasha@amazon.com>; B S, Sharath <sharab@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: Re: Retargeting metrics deck

A couple of additional points:

The M3 conversions table is updated hourly but the key for the hourly updates refer to the time at which an order occurred - not when the original click happened. The query provided for querying this table groups by the traffic event time (click). If you rerun that query a few hours apart you will observe the data changing as more orders come in associated to clicks from a given day. Will will be the case until 7 days after the click day (due to 7 day attribution window).

If querying the M3 table be sure to use the full where clause from the query below otherwise you will include traffic from SP and/or DSP etc. The query also enforces the 7 day attribution window. Also note there are two M3 tables, you want the P1D version for reporting purposes.

I recently added the campaignid field to the conversion data (in the last few days). Note that the campaignid field in M3 can be joined to campaigns but we lack the same field for click data

(so you can associate OPS but not spend). That's what we are missing from Fenix at the moment.

attributedEventsDailyManifestV2 is deprecated and is being deleted shortly - please don't start using it. We have a replacement which will generate the data faster (and will have a new different name). Note that this is basically why we're keen on decoupling you from specific table versions because we change stuff under the hood pretty frequently.

You can track the completeness of the data in Athena via

http://s9-dam-dra-build-monitoring-iad.iad.proxy.amazon.com/filesystemHealth.html#IAD/attributedEventsDailyManifestParquet.IAD/-/P14D. We automatically get a ticket when the filesystems breach the SLA.

Tony

From: Asham, Mina
Sent: 30 January 2019 10:05
To: B S, Sharath; McBryan, Tony; Balgi, Sameer; Mokarim, Zohair; Walton, Chris; Hope, Emily
Cc: Blums, Martins; Certini, Luca; Birkett, Andrew
Subject: RE: Retargeting metrics deck

Hi Sharath,

Added answers inline.

Thanks,
Mina

From: B S, Sharath <sharab@amazon.com>
Sent: 30 January 2019 03:39
To: Asham, Mina <minaasha@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Balgi, Sameer <sameerb@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Tony, Mina,

Thank you for the access and queries, this last week we were working on understanding the table schemas and columns available and we were able to generate this week WBR by ourselves. Now we have got a good understanding of data available. However, we have following questions

1) What is the frequency of data being uploaded to Athena cluster? We were running the query to get last week's data from da_athena. attributedEventsDailyManifest and data are only available till Jan 25ᵗʰ, but I was able to get the 26ᵗʰ data using da_athena. attributedEventsDailyManifestV2.
This data is updated **daily**, attributed events is always a few days behind as it relies on multiple upstream data source being available, the usual lag is ~4 days.

2) What is the type of update/insert performed these tables, is it daily insert or daily insert plus update to the existing records? For m3_conversions, da_athena.attributedEventsDailyManifest, campaign,ad, ad_decoration & decoration tables?
All data is insert only (we do not do any delta updates). For tables, m3_conversions_p1d is **hourly**, the rest are **daily** (in general tables are partitioned by their insert frequency, you can show the properties of any table and it will show the granularity).
The sp_* tables are all daily snapshots of SSPA DB, that means that each day contains all the data up to that day, so you only need the latest day. The rest of the tables you need to include all the days you want (e.g. to collect a week's worth of OPS).

3) sp_ad_decoration table only have data available before Jan 21ˢᵗ.Is it something expected or the data is missing?
Sometimes the automated Athena query loading the partitions from S3 fails. You can fix that by clicking the three dots next to the table and then "Load partitions". (I just did that for this table though).

4) While exploring the data, I stumbled upon a column **Trafficevent** in table da_athena.m3_conversions_p1d, and I found **campaignId** within the text value, is it something we can use to get the conversation data at **campaignId** level?
Yes, but this contain multiple campaign ID spaces (e.g. Fenix, Cornerstone, etc…). I think it should be okay if you limit by "ad_server = 'FENIX'".

5) One last question for now, What are **venues** column refer in the query?
Offsite remarketing campaigns are stored in the same data store for Sponsored Products campaigns, the venues is how we distinguish between onsite (aka SP) and offsite (aka Display Ads). The field is a bitarray (represented as a long). [bit0 means onsite is enabled, bit1 means offsite is enabled, bit2 means onsite for Amazon business is enabled]. So "venues & 2 = 2" reads as "extract second bit from the bit array, and make sure it's set to 1".

And we will share the final schema of data after creating couple WBRs using the current approach, so that you can pipe us the data using EDX or Andes at once with final schema.

Thanks
Sharath
From: Asham, Mina <minaasha@amazon.com>
Sent: Tuesday, January 22, 2019 10:07 AM
To: McBryan, Tony <mcbryant@amazon.co.uk>; Balgi, Sameer <sameerb@amazon.com>; B S, Sharath

<sharab@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris
<walton@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew
<abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

The advertiser name and dimensions are available now (thanks to Tony for importing the
required table), here is a sample query from our Athena tables:

```
SELECT merchant_customer_id, merchant_customer_name,
merchant_friendly_name, merchant_pg_group
FROM da_seller_central.merchant_dimension
WHERE region = 'NA' AND day = '2019-01-21';
```

-Mina

From: Asham, Mina <minaasha@amazon.com>
Sent: 22 January 2019 14:48
To: McBryan, Tony <mcbryant@amazon.co.uk>; Balgi, Sameer <sameerb@amazon.com>; B S, Sharath
<sharab@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris
<walton@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew
<abirkett@amazon.com>; Asham, Mina <minaasha@amazon.com>
Subject: RE: Retargeting metrics deck

Hi All,

I have created a couple of queries for the campaign metrics (excluding name and dimensions,
we are working on adding a dataset that includes these but we do not have it yet).
- The first query uses our Athena tables and generate all campaign metrics except the
  spend (we do not have granular spend data in our Athena tables as Tony mentioned).
- The second query uses ACH (SSPA Redshift cluster, link included in file), and covers the
  same plus the spend

A few things to note:
- The campaign model metrics (campaign metrics without spend), are a daily snapshot,
  so the values could change within the day (e.g. budget can be changed multiple times
  but the snapshot will have the last value)
- The campaign model is (simplified): campaign -> ad group -> ad, each of them have its
  own status (status is aggregated top to bottom), my query only returns the campaign
  status which is the one advertisers usually toggle more than the others
- Advertiser name and dimensions should be available soon, will update when we have
  them

Thanks,
Mina

From: McBryan, Tony <mcbryant@amazon.co.uk>
Sent: 22 January 2019 14:32
To: Balgi, Sameer <sameerb@amazon.com>; B S, Sharath <sharab@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; Asham, Mina <minaasha@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

Hiya,

I don't think we have any of that. At least not at the moment.

1) Order ID/Campaign ID – We have a daily dump of the campaign setup (that's the stuff Mina is working on for you) but Fenix doesn't vend the campaign ID's it to us at event time so we can't currently join any performance data against it.
2) Budget – They do enter a budget (available in the daily dump) but it's not clear how you'd present this (the objective isn't the same as DSP to exactly spend all the budget, they are guardrails instead). Again we can't currently join performance data against campaign data.
3) Goal – I'm not aware of the ability to specify a goal other than CPC bid amount.
4) Billing – We don't do the billing, Iron Mountain do that for us. Not sure what you'd extract?

I'd like us to get the ability to join campaign data against performance data (it's been a pain point before) but I think it's pending us moving away from Fenix so we can get the join key. That'd get us budget utilization as well (although I'm less convinced this is useful because it's not the same as a DSP budget).

Tony

From: Balgi, Sameer <sameerb@amazon.com>
Sent: 22 January 2019 14:21
To: McBryan, Tony <mcbryant@amazon.co.uk>; B S, Sharath <sharab@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; Asham, Mina <minaasha@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

Hi Tony,

I'll let Sharath provide inputs on the technical front, but thinking of the metrics deck will be helpful to have the below. If these are already covered, that's great.

I) Order ID or Campaign ID – Can we bring in data at a campaign ID level? Does this blow up the data-set significantly?
II) Budget – Do advertisers enter a budget in the UI? Can we export this data point? This can live in a separate table – but if we can bring in something that tells us the budget amount by advertiser ID, would be great to see how much budget we utilized by advertiser?

III)   Goal – Are we getting inputs from Advertisers on the goals they enter for a campaign in the UI? Would be helpful to get that as we can then compare how we did vs. goal.

IV)   Billing – do we save Billing data in your cluster? Not an immediate priority, but if your team houses the billing info, we should set up the data pipeline to extract that.

Thanks for all your help in setting these up.

Thanks,
- Sameer

From: McBryan, Tony
Sent: Tuesday, January 22, 2019 3:20 AM
To: B S, Sharath <sharab@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; Asham, Mina <minaasha@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Balgi, Sameer <sameerb@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

Hiya,

For a long term solution we'd rather vend a file with a fixed schema which we can vend with agreed SLA rather than the files we currently load into Athena. This means we can guarantee that the column names, semantics etc won't change under your feet.

Do you have opinions on the semantics of the file(s) you'd want?

I imagine we'd vend two files, one with the set of attributed orders that happened on that day (the MADS M3 conversions dataset) and one with the set of impressions/clicks that happened that day (ads dataset).

Options are:

1)   Group/feature fields. I'd suggest **click_day** (YYYY-MM-DD UTC), Amazon **marketplace_id** (i.e. US = 1, UK = 3) and **advertiser_id** (numeric) with an additional **order_day** (YYYY-MM-DD UTC) field in the conversions
2)   Measure fields. I'd suggest for the ads dataset to have **impressions** (numeric count), **clicks** (numeric count), **ad_revenue** (sum in microdollars, always USD), **media_spend** (sum in microdollars, always USD) and for the conversions dataset to have distinct_conversions (numeric) and ops (sum in microdollars, always USD).

Do you have a preference for delivery mechanism? We can do S3 as you mentioned below formatting in either CSV, TSV, JSON or Parquet. Note that as HC data we typically client side encrypt our data files using Odin or KMS as key store. Alternatively we can vend via one of the standard Amazon mechanisms like EDX or Andes which I think we'd probably prefer in the long run.

SLA-wise we can probably pin down once we've got the working implementation but I'd imagine it's going to be a bit over 1 day (from end of day) for conversion data and a bit over 2-3 days for ad data.

Tony

Hi Tony,

Thank you for the details.

For long term solution, you can provide access to the S3 bucket and I can copy the required data into our redshift cluster. How to grant access steps: https://stackoverflow.com/questions/36730820/copy-to-redshift-from-another-accounts-s3-bucket
Our AWS account and role : arn:aws:iam::**006877342108**:role/Redshiftunloadcopy. I testing around at Athena cluster , will get back if I have any questions on it.

Thanks
Sharath

I've added you to the Dynamic Ads team so you should be able to access the console via https://access.amazon.com/aws/accounts/fetchConsoleUrl?account_name=Dynamic+Retail+Ads&session_duration=43200

I'm not sure this is the long term best way of giving you access but it'll get things going for now. I think in the long run we'd probably want to create IAM roles and assign it that way

The queries we run to generate the WBR numbers today are all in https://code.amazon.com/packages/DraMLBiddingSandbox/blobs/mainline/--/Tools/wbrMetrics.py

Specifically you can run the queries below in Athena:

MADS (M3) conversions (i.e. 7 day attributed OPS) by day:

```
WITH
```

```
    m3 AS (
      SELECT
        *,
        date_parse(date_time, '%Y-%m-%d %H:%i:%S:%f') as date_time_d,
        date_parse(trafficEventDateTime, '%Y-%m-%d %H:%i:%S:%f') as
trafficEventDateTime_d
      FROM da_athena.m3_conversions_p1d
      WHERE region='NA' AND day>'{0}'
    )
    SELECT
      date_format(trafficEventDateTime_d, '%Y-%m-%d') AS day,
      SUM(ops) / 1000000.0 AS offsite_sales,
      COUNT(DISTINCT concat(splogid,spqualifier)) AS
offsite_conversions
    FROM m3
    WHERE date_diff('second', trafficEventDateTime_d, date_time_d) <
60 * 60 * 24 * 7 AND ad_server = 'FENIX'
    AND dimAlgorithmId IN (7, 37) AND slot_name LIKE 'sp_syn_da_%'
    GROUP BY 1
```

Impressions, clicks, ad revenue and media spend (by day):

```
    SELECT
      day,
      SUM(impressions) AS ad_impressions,
      SUM(CAST(spClicks > 0 AS BIGINT)) AS clicks,
      SUM(spCpc / spClicks) / 1000000.0  AS ad_revenue,
      SUM(winningPrice) / 1000000000.0 AS spend
    -- You could also use attributedEventsDailyManifestV2 which is one
day earlier. Only has data since 2018-12-28 though.
    FROM da_athena.attributedEventsDailyManifest
    WHERE region='NA' AND day>'{0}' AND spLogId IS NOT NULL
    GROUP BY day
    ORDER BY day
```

Both of these should also be groupable by advertiserid if you need to.
For the m3_conversions_p1d table there is an advertiser_id column already.
For the attributedEventsDailyManifest table you can get advertiserid with a select of column
"cast(json_extract_scalar(rawsourceevent, '$.spAdvertiserId') as
bigint)".

The remaining columns are trivially computable from the data above (i.e. CTR = # clicks / # impressions).

We also read counts of advertisers from the SP ACH table (I can't give you access to this as it's not our cluster). See http://sspa-db-accounts.aka.amazon.com/ for details.

```
SELECT COUNT(DISTINCT entity) FROM bi.advertiser_on_site_perf
&
SELECT COUNT(DISTINCT advertiser_id) FROM
bi.advertiser_syndication_beta
```

The latter two ACH tables are not backfillable, you can only get a point in time measurement for the current value so you have to do the query at roughly the same point in time.

Tony

From: B S, Sharath <sharab@amazon.com>
Sent: 20 January 2019 20:28
To: Mokarim, Zohair <zohaim@amazon.com>; Walton, Chris <walton@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; Asham, Mina <minaasha@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>; Balgi, Sameer <sameerb@amazon.com>
Subject: RE: Retargeting metrics deck

+Sameer for being in the loop.

From: Mokarim, Zohair <zohaim@amazon.com>
Sent: Saturday, January 19, 2019 9:36 PM
To: Walton, Chris <walton@amazon.com>; McBryan, Tony <mcbryant@amazon.co.uk>; B S, Sharath <sharab@amazon.com>; Asham, Mina <minaasha@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>
Cc: Blums, Martins <martinsb@amazon.com>; Certini, Luca <certinil@amazon.com>
Subject: RE: Retargeting metrics deck

-Moving non action item owners to BCC

Hi Chris/Tony

Thanks for meeting last week. I think we should be able to get some quick wins if we are able to get access to the DA Athena cluster as soon as possible to get data at the lowest grain and hopefully can further develop some of the reporting which Mina currently has. Here is a summary of the data points we discussed along with proposed owners and next steps to move this forward. I have put in some initial ETA's which we can firm up as each work stream progresses.

Performance Metrics (Advertiser Level)	Owner	Joint Owner	ETA	
Active Advertisers Ad Revenue Cost of Revenue (Traffic Acquisition Cost) Gross Margin Impressions (MM) CTR CVR CPC ACOS / ROAS Clicks Avg Asin Price SP Attributed OPS (7 day)	mcbryant@	sharab@	1/25/2019	1. Data is available in the DA Athena C[...] 2. Tony to provide access to Sharath (Sh[...]
Comparable SP metrics for active advertisers	sharab@	zohaim@	2/1/2019	1. Join with SP datasets once above dat[...]
Experimentation Data (SIMS?)	walton@		1/25/2019	
Detail Page Views (Vendor/Seller) Distribution % of DPV Remarketed	sharab@	zohaim@	1/25/2019	1. Dependence on ASIN data below
OPS Benefit From Remarketing	zohaim@	certinil@	1/25/2019	1. Align on methodology with Oksana
Campaign Metrics	**Owner**	**Joint Owner**	**ETA**	
Advertiser Name Advertiser Dimensions - Decile, Consumer Rollup etc Number of Active Campaigns / Budget / Utilized Number of Campaigns Created ASIN's in Ad Campaign Advertiser Retention Campaign Status (Paused/Archived)	minaasha@	certinil@ zohaim@	TBD post call with mina	1. Zohair & Luca to setup a call with Min[...]
UCB - Advertiser Experience	ehhope@	zohaim@ certinil@	TBD	1. Emily to help combine efforts on this
Bidding Metrics	**Owner**	**Joint Owner**	**ETA**	
Model Accuracy Slides (Chris) Bid Utilization - Charged Bid / CPC Internal / External Auction Metrics Remarketing Funnel Analysis (MBR)	walton@		1/25/2019	1. Putting Chris down as an initial place[...]
Supply Metrics (By Source)	**Owner**	**Joint Owner**	**ETA**	
Bid Requests Bids Placed Win Rates	zohaim@	A9 (Pavel Kalinin)	Q1 2019	1. Zohair to setup time with Chris / Dre[...]

Regards
Zohair

From: Walton, Chris <walton@amazon.com>
Sent: Thursday, January 17, 2019 11:51 AM
To: Zbagerska, Ruslana <ruslana@amazon.com>; Blums, Martins <martinsb@amazon.com>; Mokarim, Zohair <zohaim@amazon.com>; Hope, Emily <ehhope@amazon.co.uk>; Anderson, Keith <keithand@amazon.com>; Deng, Emily <emilyd@amazon.com>; Abdelmonem, Yousef <yabdelmo@amazon.com>; Park, Eileen <ejpark@amazon.com>
Cc: Bradley, Drew <drewb@amazon.com>; Birkett, Andrew <abirkett@amazon.com>
Subject: RE: Retargeting metrics deck

Yes – I am happy for my team to assist while we hire a BIE.   We had a meeting this morning and identified the main sources of data to improve the deck.

Chris

FYI. Chris as we discussed yesterday, it would be great if you and your team can pitch in short term while we are hiring the BIE.

Thank you!
Ruslana

Paul,

We will redouble our efforts on building out our databases to be able to provide more meaningful metrics.

And we will share with you the specific data points you requested in the meantime.

Martins

Thanks, Martins.

This isn't much further developed since I last asked about this. I view these decks as operating tools, particularly important for a nascent program like retargeting.

In the absence of a meaningful metrics deck, I'd really like to see answers to the items that I asked about in the meeting today, including detail page view volumes by vendor and head 3P sellers. Systematic tracking of Amazon OPS benefit from retargeting (aligned with DSP measurement).

Paul

From: Blums, Martins
Sent: Tuesday, January 15, 2019 2:26 PM
To: Kotas, Paul <kotas@amazon.com>
Cc: Zbagerska, Ruslana <ruslana@amazon.com>
Subject: RE: Retargeting metrics deck

+ Ruslana

Paul,

Thanks for the email. Please find the SDP WBR for Wk 2 that we just published.

You'll notice that we have pages that are not yet filled out or have certain data missing. The deck is not yet up to the standard we have for standard WBRs. The data infrastructure we rely on is very fragmented, at times is unavailable, and not every metric that we think important is captured. This week, while the EDI team is here in person, we have a number of sessions to align on the missing pieces and timelines of delivery. So I expect we will be able to show progress, but we are not there yet.

Please let us know if you have any questions.

Thanks,

Martins

From: Kotas, Paul <kotas@amazon.com>
Sent: Tuesday, January 15, 2019 12:54 PM
To: Blums, Martins <martinsb@amazon.com>
Subject: Retargeting metrics deck

Could you please send me the latest deck? Thanks.

paul