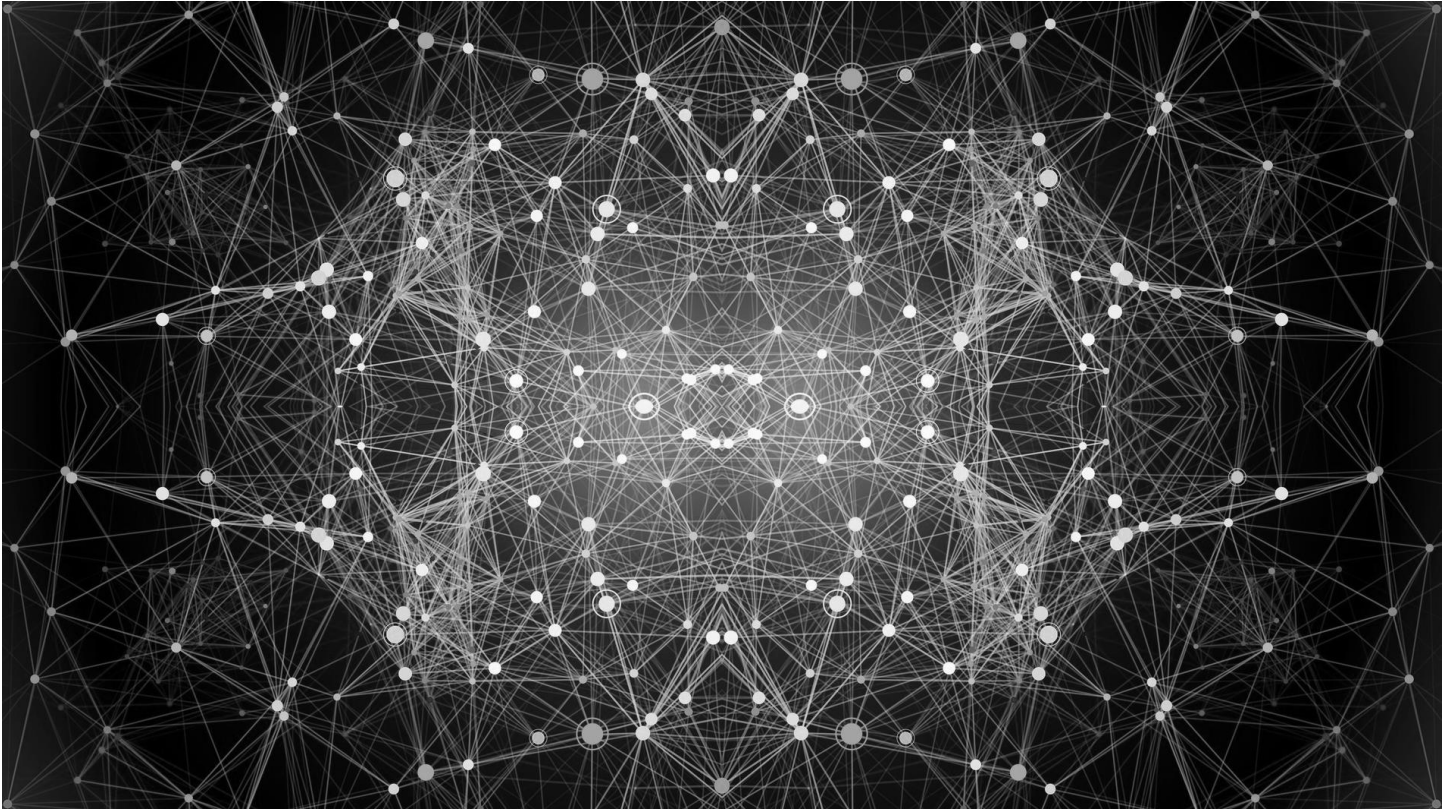

Attribution Modeling From CLD



Essentia Attribution Modeling
Whitepaper

AuriQ Systems Inc.
Pasadena, CA



Overview

Essentia is a big log processing service that AuriQ Systems has been providing for customers of Sizmek (formerly known as MediaMind). A much requested type of big data analysis from Sizmek's clients is the ability to apply attribution models against the cookie level data (CLD) logs from Sizmek's servers. In partnership with Sizmek, AuriQ has developed a process to fulfill the needs of customers for attribution analysis using current CLD logs. This document outlines the requirements and flow for this process.

Requirements

1. CLD in zipped files archived in S3

Sizmek makes available to their clients various CLD files in zipped format which can be downloaded from their ftp servers. The types of CLD log files include:

- *Standard: impression & click events log*

| | | |
|---|-----|---------|
| MM_CLD_Standard_Agency_74825_Daily_140423.zip | Apr | 5.24 GB |
| MM_CLD_Standard_Agency_74825_Daily_140422.zip | Apr | 5.23 GB |
| MM_CLD_Standard_Agency_74825_Daily_140421.zip | Apr | 5.10 GB |
| MM_CLD_Standard_Agency_74825_Daily_140424.zip | Apr | 5.09 GB |
| MM_CLD_Standard_Agency_74825_Daily_140425.zip | Apr | 4.55 GB |
| MM_CLD_Standard_Agency_74825_Daily_140426.zip | Apr | 3.50 GB |
| MM_CLD_Standard_Agency_74826_Daily_140410.zip | Apr | 3.17 GB |
| MM_CLD_Standard_Agency_74829_Daily_140428.zip | Apr | 3.09 GB |
| MM_CLD_Standard_Agency_74826_Daily_140409.zip | Apr | 3.07 GB |

- *Rich: rich media impression events log*

| | | |
|---|-----|----------|
| MM_CLD_Rich_Agency_74825_Daily_140414.zip | Apr | 842.9 MB |
| MM_CLD_Rich_Agency_74825_Daily_140418.zip | Apr | 747.2 MB |
| MM_CLD_Rich_Agency_74825_Daily_140416.zip | Apr | 739.7 MB |
| MM_CLD_Rich_Agency_74825_Daily_140417.zip | Apr | 739.6 MB |
| MM_CLD_Rich_Agency_74825_Daily_140415.zip | Apr | 733.3 MB |
| MM_CLD_Rich_Agency_74825_Daily_140419.zip | Apr | 715.7 MB |
| MM_CLD_Rich_Agency_74829_Daily_140414.zip | Apr | 674.0 MB |
| MM_CLD_Rich_Agency_74829_Daily_140418.zip | Apr | 557.5 MB |
| MM_CLD_Rich_Agency_74829_Daily_140419.zip | Apr | 547.4 MB |
| MM_CLD_Rich_Agency_74829_Daily_140417.zip | Apr | 546.2 MB |
| MM_CLD_Rich_Agency_74829_Daily_140416.zip | Apr | 544.6 MB |

- *Conversion: conversion events log*

| | | |
|---|-----|--------|
| MM_CLD_Conversion_Agency_74825_Daily_140422.zip | Apr | 1.2 MB |
| MM_CLD_Conversion_Agency_74825_Daily_140423.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74832_Daily_140425.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74825_Daily_140425.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74832_Daily_140407.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74825_Daily_140424.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74832_Daily_140408.zip | Apr | 1.1 MB |
| MM_CLD_Conversion_Agency_74832_Daily_140410.zip | Apr | 1.0 MB |
| MM_CLD_Conversion_Agency_74825_Daily_140421.zip | Apr | 1.0 MB |
| MM_CLD_Conversion_Agency_74832_Daily_140411.zip | Apr | 1.0 MB |

- *Match: master tables*

| | | |
|--|-----|--------|
| MM_CLD_Match_Agency_74828_Daily_140404.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140414.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140405.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140407.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140415.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140416.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140403.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140424.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140411.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140410.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140413.csv.zip | Apr | 1.2 MB |
| MM_CLD_Match_Agency_74828_Daily_140429.csv.zip | May | 1.2 MB |

All files are downloaded as is from **ftp.mediamind.com** and stored in Amazon S3. S3 storage is included with the service, but if the client wants to use their own S3 bucket, then this can be easily configured.

2. Essentia account

To get an Essentia account, simply access **http://essentia.auriq.com** from your browser and follow the sign-up procedure.

AURIQ ESSENTIA

Login

Email Address

Password

[Forgot password?](#)

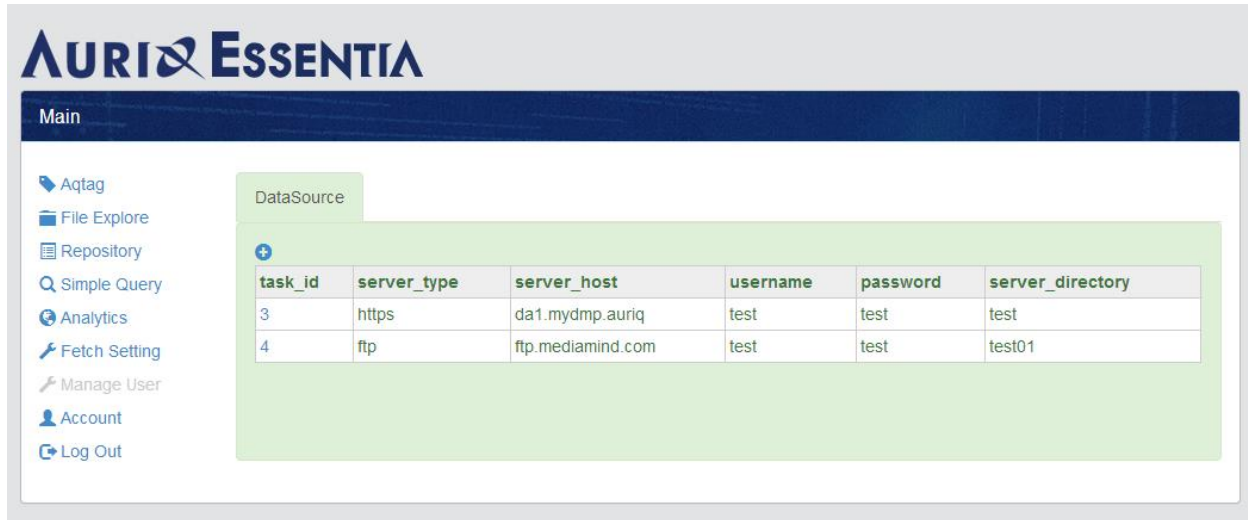
☐ Remember me

Login Use my Google account

Don't have an Account yet? [Sign Up Here](#)

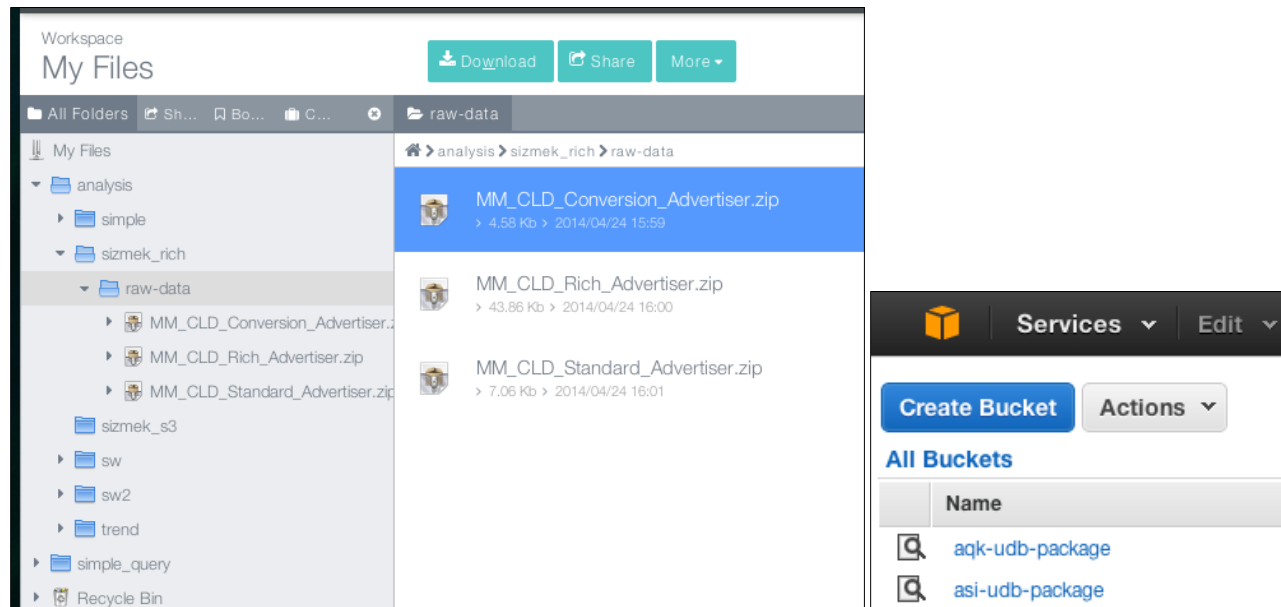
Process Flow

- 1) From the Essentia user interface, you specify which server(s) you want to fetch data from.



(Note: By default, Essentia will fetch CLD files daily. If it is necessary to fetch data more frequently, then this can be modified by AuriQ)

- 2) CLD files are stored in S3



3) Choose which attribution models you want to apply to your CLD.

Models

+

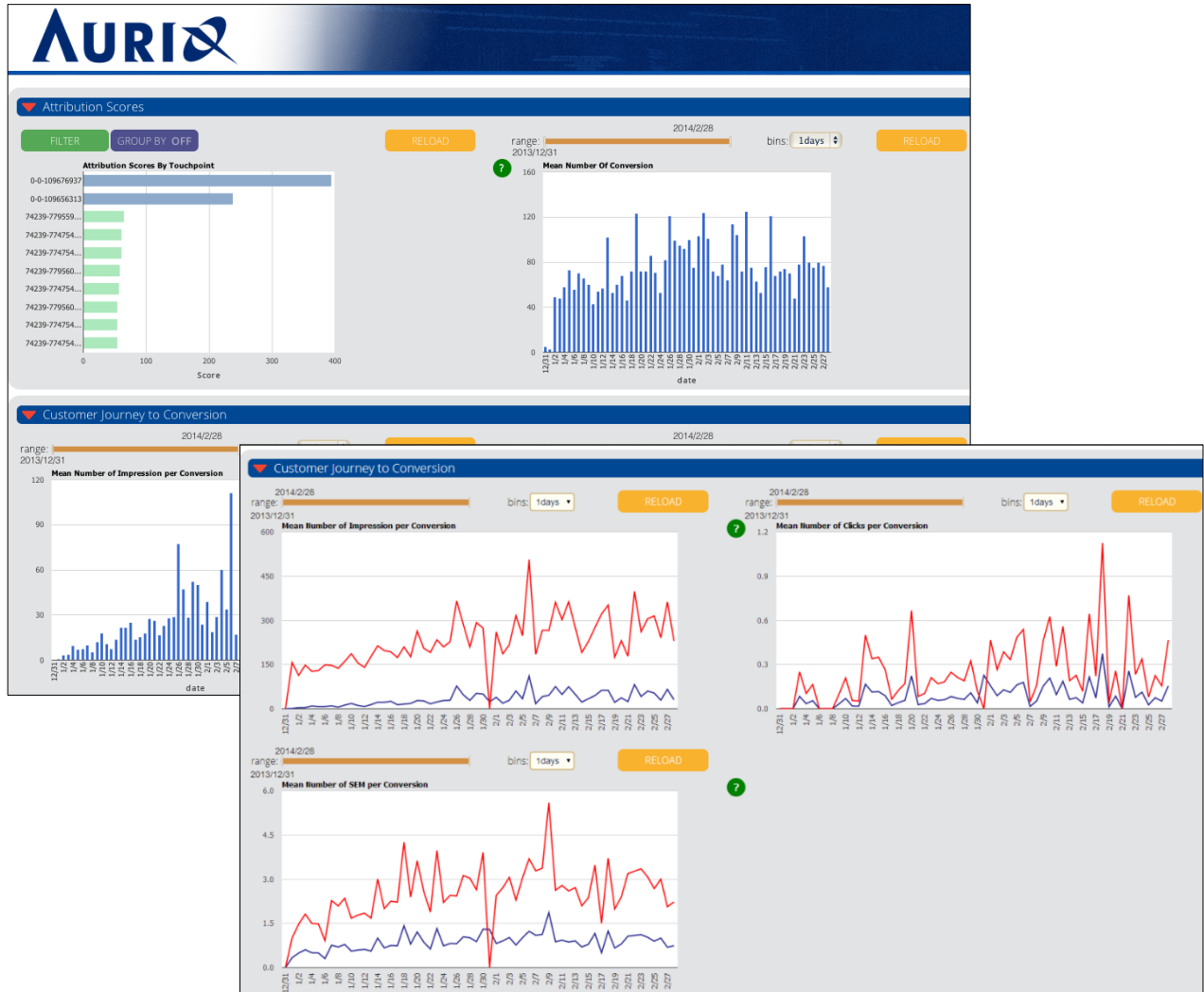
| | Name | Description | |
|-------------------------------------|--------------------------------|---|------|
| <input type="checkbox"/> | Last Click Only | Account only last click before conversion | Edit |
| <input type="checkbox"/> | First Click Only | Account only first click to conversion | Edit |
| <input checked="" type="checkbox"/> | All Click Even Distribution | Account all click equally | Edit |
| <input checked="" type="checkbox"/> | All Click Time Decay | Account all click with time decay weighted | Edit |
| <input type="checkbox"/> | Imp & Click Weighted | All imp and click weighted by a fixed ratio | Edit |
| <input type="checkbox"/> | Imp & Click CTR | All imp and click. Imp weighted by CTR | Edit |
| <input checked="" type="checkbox"/> | Imp & Click & Organic Weighted | Also include Organic search | Edit |
| <input type="checkbox"/> | Custom | Support any custom algorithm | Edit |

Run

4) Set or modify attribution parameters

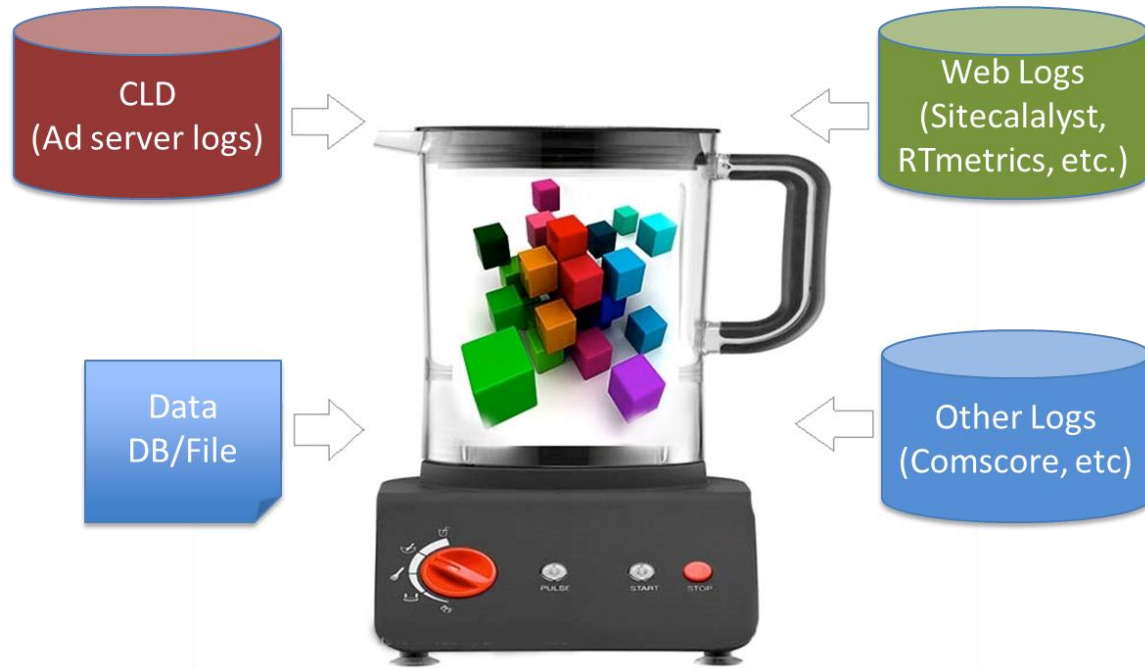
```
# Specify the time range over which conversion events
# will be identified (format is : YYYY-MM-DD)
conversion_start_date=2014-01-01
conversion_end_date=2014-01-10
# Relative weight of impressions (MM, EventTypeID==1)
w_imp=1.0
# Relative weight of clicks (MM, EventTypeID==2)
w_click=1.0
# Relative weight of organic searches (As found in AQLogs)
w_seo=1.0
# Relative weight of search clicks (SEM) (MM, EventTypeID==3)
w_sem=1.0
# Weigh impressions by the ratio of : #clicks/#impressions for a given SiteID
# (if 0, no weighting, if 1 weighting)
weigh_imp_by_ctr=1
# half-life (in sec) of the time_decay attribution model
half_life=43200
# time binning (in sec) use to write the distribution of events prior to
# conversion.
tbin=3600
# Attribution model used. Models currently available include :
# linear, first_click, last_click, first_interaction, last_interaction, and time_decay
model=linear
# Number of conversion tag IDs of interest for the attribution calculation :
nb_conv_tag=9
```


- 5) Run the analysis and view output.



Data Enrichment

The process outlined thus far extends only to attribution based on CLD only. To better understand the customer journey, it is possible to enrich the analysis by incorporating other types of data.





Want to learn more about AuriQ
Essentia's big log processing
capabilities? Do you have a project
in mind already?

Please contact us for a free consultation. We'll
assess your data analysis needs and help you
determine if AuriQ Essentia is right for you.



AuriQ Systems, Inc.

301 North Lake Ave, Suite 1001

Pasadena, California 91101

(626) 564-2781 essentia@auriq.com