

Entire report date range DD/MM/YYY - DD/MM/YYYY

Currently, this is the smallest unit of measurement we are interested in

For a particular DD/MM/YYYY

All rows are processed each of N passes

N passes on fields grouped by dependencies

Key

Pass 1

Bound

Pass 2

Distributed

Aggregate

Pass 3

Derived

Text

impressions	webLeads	phoneLeads	sum(impressions)	phoneLeads+webLeads
100	20	3	500	23

impressions	webLeads	phoneLeads	sum(impressions)	phoneLeads+webLeads
200	40	7	500	47

sum(impressions)	300
sum(webLeads)	60

This formula is built-into the system for distributed fields; dependant field is user set.

$$\frac{\text{impressions}}{\text{sum}(\text{impressions})} * \text{phoneLeads}$$

(A, Y)

(A, X)

(B, Y)

(B, X)

sum(impressions)	200
sum(webLeads)	50

Bound / distributed fields first

Aggregate / derived fields always after

normal aggregates
reference entire
report range

sum(impressions)	500
------------------	-----

sum(webLeads)	60
---------------	----

sum(phoneLeads)	30
-----------------	----

Parent key: (A, _)

sum(impressions)	300
------------------	-----

sum(webLeads)	60
---------------	----

phoneLeads(impressions)	10
-------------------------	----

Parent key: (B, _)

sum(impressions)	200
------------------	-----

sum(webLeads)	50
---------------	----

phoneLeads(impressions)	20
-------------------------	----

In this example 3 passes are performed over the data. In actuality, the number of passes is determined based on the dependencies between the fields (references within formulae).

Pass 1: all bound fields (i.e. fields without dependancies) processed, internal summations for these fields are also created in case they are referenced later on.

Pass 2: any distributed fields (i.e. fields whose attributes from which they derive are less granular) are processed; these are attributes for which we do not have specific row values, only more generalised values that span several rows (achieved by way of different data sources returning a multipart key); these must be distributed only for a specific date. We only allow bound fields as the proportion source for the distributed fields (in this case “impressions”) because it allows us to distribute the field values early on, while we are still only working with attributes from data sources.

This creates a separation from attribute sourced fields (passes 1 & 2) and formulae derived fields (pass 3). From pass 3 onwards, all attribute values have now been filled in, and only derived fields need be handled.

Pass 3: all passes after the first two are actually derived fields which pass through the formulae evaluator; these also produce summations as a result of their evaluation which can be used in footers.

