

allow group by (or raw sql) to define dynamic sets with more general conditions (Released in 9.2)

Executive Summary

Provide a way to define a set based on a condition that involves multiple rows in a database table. Currently, sets can be defined using conditions or formulas that test each data row individually. This makes it apparently impossible to define a set such as "Most recent Order Id for each product"

NOTE - Although this idea has been marked released, that was because LOD calculations can achieve at least the main use case described here via a very different approach and query. There may still be value in considering the idea as a way to generate correlated subqueries, and (I think) be a more natural solution to some use cases.

The problem this idea solves

Frequently, customers have a long stream of records and would like to see information about the latest record in each category, e.g., show the amount of the most recent sale in each category, the status flag in the most recent status history record for each monitored device, etc. Besides my own customer real world examples, there were at least two requests to the forums about this in just the last few weeks, and another today on stack overflow. So it is a common use case.

Current workarounds

There are currently two ways to achieve this in Tableau, but both impose unnecessary costs, and I think there may be a way to generalize the condition tabs on sets to provide a better solution.

First, the currently available workarounds:

1. Make a custom sql data connection to only include the latest record in each group

for example, see the subquery in the where clause below

```
select [Order ID], Item, Sales, [Order Date] from Superstore
where [Order ID] in (select max([Order ID]) from Superstore group by Item)
```

2. Use table calculations to partition the records ordered by date and only display the last one in each partition

The table calc has the disadvantage of requiring sending all the data across the wire to the client, merely to throw most of it away. Not good for performance. The custom sql approach has the disadvantage of complexity and also Tableau can't optimize custom sql well, such as by not selecting for unused columns, leaving you at the mercy of a good query optimizer.

Proposed solutions

allow group by (or raw sql) to define dynamic sets with more general conditions (Released in 9.2)

My suggestion is to expand the condition tab functionality for sets to allow you to specify bolded subquery above. I think of sets as a way of defining what goes to the right of an "in" keyword in SQL. So either:

- add a group by option to the condition and/or top tabs so we can say, for instance,
 - include the top 1 by max([Order Date] group by Item
- and/or, let us provide raw sql in an advanced tab for a set where we can provide a select statement to define the set, such as the bolded subquery in the above example

If anyone knows a better workaround, I'd love to hear it. For now, I'll stick with making views similar to the custom sql above.

Note, I know you can easily ask for the max([Order Date]) for each item. The difficulty comes when you want to efficiently also, display the other information from that last order.