scenario

Gold layer table

Silver layer table

Bronze layer table

semantic layer table

Tableau

DQ score =70 DQ score =100 DQ score =90 DQ score = 80

Bronze layer has raw data.

let’s say we have an adaptive rule on top of this and it gives DQ score of = 70 (may be because of nulls,duplications etc.)

And silver layer has the transformed data.

nulls are removed, dupes are removed .(Basically cleaned data)

let’s say we have another adaptive rule here and it gives DQ score of =100 (because there is no nulls,duplications etc.)

And gold layer has business aggregated data.

let’s say we have a custom rule on top this to check values based on a certain condition.

and it has given a DQ score of = 90 (because some rows are breaking )

And in semantic layer let’s say we have a view for a Tableau dashboard which itself has a DQ score of = 80

what to be solved?

we want to give each Tableau dashboard an aggregated Data Quality score considering this table lineage

How can we represent the DQ score based on downstream layers? <- an aggregated score

as an example in the above scenario

let’s say we have defined our logic as AVG(downstream DQ Scores)

so -> Dashboard level aggregated DQ score = (70+100+90+80)/4

Note: we have integrated DQ jobs to Collibra DIP and this final avg DQ score output should come from Collibra DIP.

1. what aggregations /steps we have to do to achieve this, or is this already available ?

existing aggregations DQ-DIP aggregations

