

Task 1

1. Some of the problems with SQL and data analysis is that SQL Group By method used in data analysis does not allow construction of histograms. As the analyst most likely would like to group by time, or location. It could be done indirectly but would take a lot of experience with SQL and nested queries. A bigger problem is the lack roll-up and drill-down problem. Aggregated sublevels of data would result in many empty cells as not all cells can form keys. However aggregation of columns for roll-up of N elements would result in creating 2^N new columns.

2. Time: Hour < Day < Week < Month < Year < Decade < All

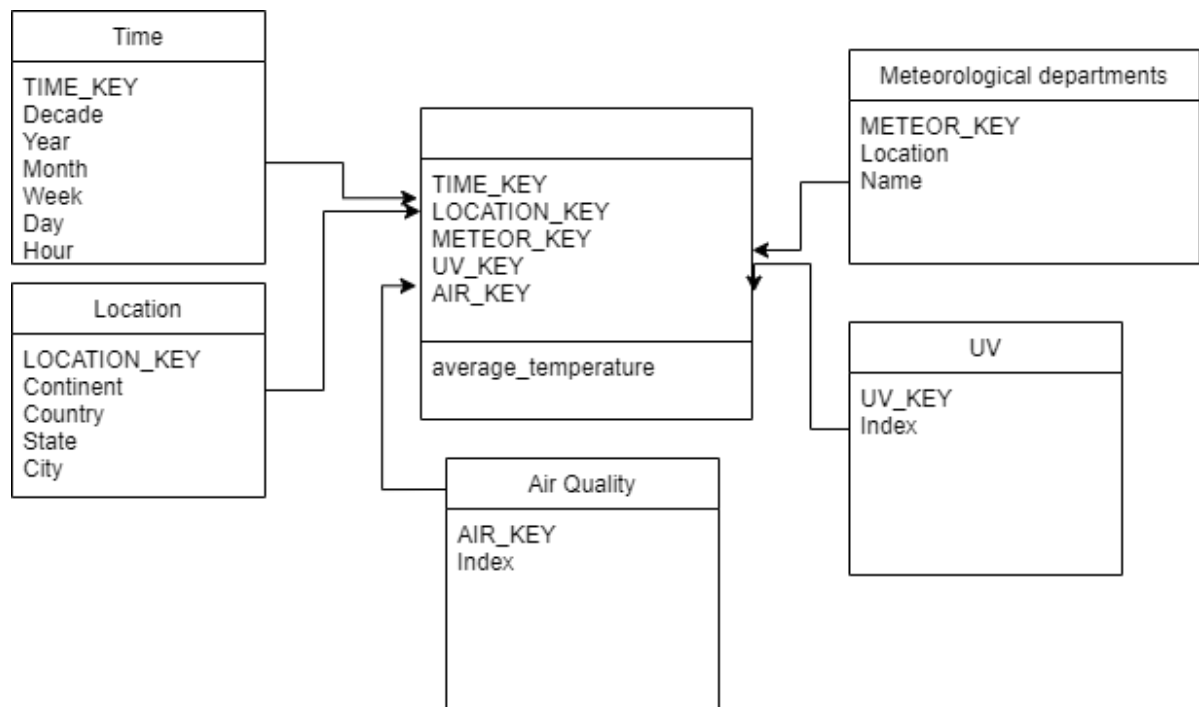
Location: City < State < County < Continent < All

Meteorological departments: Name < Location < All

UV: [1, 11] < All

Air quality: [0, 201] < All

For air and UV would use intervals for better concepts.



- 3.

$$4. (L_{time} + 1) + (L_{location} + 1) + (L_{metdep} + 1) + (L_{UV} + 1) + (L_{Air} + 1) = (6 + 1) + (4 + 1) + (2 + 1) + (1 + 1) + (1 + 1) = 19$$

We would need 19 cuboids for full materialization

5. Roll up location: City → Continent

Roll up Time: Hour → Year

Slice: Continent = "Europe" AND Year="This year"

Roll up Time: Year → Decade

Slice Continent = "Europe" AND (Decade="Last")