There have been a lot of cool new projects within the industry focused on parsing 'meaningful data' from blockchains (the most useful currently is the blockchain-etl project imho: https://github.com/blockchain-etl/ethereum-etl). However, it's clear that the architecture of ETH2.0 is noticeably different from that of ETH1.0; how will this likely affect data analysis projects like blockchain-etl?

Most data analysis of this ilk requires one to parse through the whole state of a given blockchain from genesis and then transform that data into a more convenient format. Given that there could be around 1,024 shards in ETH1.0 plus the beacon chain AND I remember reading that the legacy chain could – in theory – exist on a single shard; it seems inconceivable that a single data analyst could run all shards in order to have access to all the data ETH2.0 may bring, lest we allow all ETH2.0 data analysis to happen on Google BigQuery.

Questions:

Is this problem something that anyone has thought about?

Is this actually an important concern or am I mistaken? If so, why?

Thanks!