

Water Cooler July-14 2022

Notes/Observations by James Madison from the open discussion after the TSC call on July-14 begin below.

After the business meeting, we had a "hallway conversation" to discuss a specific aspect of HDS design.

The current stat model is much loved by those who use it. In particular, the orientation around stat plans, then premiums and losses, then the grain of policy, object, subline, coverage, class, transaction code--this pattern captures the essence of the nature of insurance, and can be understood and supported by many folks, particularly those without a detailed data design skills.

We all worry a bit that it's stat plan oriented. But it shouldn't be a show stopper. So that aside, the premium/loss nature and the nature of the grain is largely agnostic to stat plans. It is just insurance. And there is no reason to believe that nature will change.

But we do want to put dozens and dozens of new attributes on this basic skeleton, so let's do that.

Thus, we want to make that flat model the base of the HDS. This leads to:

- **Agreement #1: Let's use the stat model but with lots more attribution as the base of the HDS.**

This immediately produces a challenge--flat models can be brutal to query. This leads to what James humorously (but not kidding) calls "Doing ETL in the BI layer". That is, you often have to decompose the data, mostly through aggregates, group by's, and lots of filters, into some other form before doing meaningful.

As a result, it's common to introduce a more robust model that does actual ETL, instead of stuffing it into your reporting/BI layer. However, we don't know that the nature of data calls will cause sufficient pain to worry. So:

- **Agreement #2: Maybe stat and data calls aren't that hard to do from a flat model, and we're good, but we don't know.**

And we don't know. This is EXACTLY the work Peter is doing--is it a bread box or elephant, that query mess after the flat tables but before the reports. Don't know. Let's do a few and see. And that is in motion.

However...it seems SOMETHING will have to be done after the flat tables and before the reports...to that we said, whatever that is:

- **Agreement #3: Decouple that mystery layer from the flat layers.**

This is the big agreement. Just from engineering 101, if we separate them, we have now freed the flat model that we all pretty much think we need, to run fast, and then Day 2, we make this mystery layer, whatever it is. Could be virtual, could be the most beastly DW you ever saw. Whatever it is, we don't mess with the basic flat tables on the left side of the flow that all the business people love, and that even a small carrier might have a hope of building.

This leads to the question of ownership of the mystery layer. Would it have to be the carrier? Is it the analytics node? Is it some new party? If it's a new party, that could be an entire business model. But if we do put it in the HDS proper, it would put the load on the carriers, which is hard for a number of reasons, which we won't explore here for time. But this leads to:

- **Open question #1: Who owns the mystery layer? Don't assume it's the carrier. Stat agencies? Still undefined.**

Key take away though: we seem to love the base of the HDS being the current flat model, just with lots and lots of attributes. So let's lock that down so we can run that track fast, then we can take up the other parts in order of value.