

2022-9-09 Meeting Agenda



GMT20220909-1...1920x1080.mp4

Date

09 Sept 2022

Antitrust Policy Notice

Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Updegrove LLP, which provides legal counsel to the Linux Foundation.



This is a weekly series for The Regulatory Reporting Data Model Working Group. The RRDMWG is a collaborative group of insurers, regulators and other insurance industry innovators dedicated to the development of data models that will support regulatory reporting through an openIDL node. The data models to be developed will reflect a greater synchronization of data for insurer statistical and financial data and a consistent methodology that insurers and regulators can leverage to modernize the data reporting environment. The models developed will be reported to the Regulatory Reporting Steering Committee for approval for publication as an open-source data model.

openIDL Community is inviting you to a scheduled Zoom meeting.

Join Zoom Meeting

<https://zoom.us/j/98908804279?pwd=Q1FGcFhUQk5RMkEpkVlFTWtXb09jQT09>

Meeting ID: 989 0880 4279

Passcode: 740215

One tap mobile

+16699006833,,98908804279# US (San Jose)

+12532158782,,98908804279# US (Tacoma)

Dial by your location

+1 669 900 6833 US (San Jose)

+1 253 215 8782 US (Tacoma)

+1 346 248 7799 US (Houston)

+1 929 205 6099 US (New York)

+1 301 715 8592 US (Washington DC)

+1 312 626 6799 US (Chicago)

888 788 0099 US Toll-free

877 853 5247 US Toll-free

Meeting ID: 989 0880 4279

Find your local number: <https://zoom.us/j/98908804279>

Attendees

Peter Antley (AAIS)

Dale Harris (Travelers)

Ash Naik

Birny Birnbaum

BourjailiHi

Brian Hoffman

James Madison

Jeff Braswell

Ken Sayers

Libby Crews

Lori Dreaver Munn

Mike Nurse

Patti

Sandra Darby

Reggie Scarpa

Brian Hoffman

Agenda

- a. AAIS progress on Auditing the math
 - i. Whats been done
 - ii. Whats been left to do
- b. Review our plans for Homeowners
- c. Evaluate Health of Group overall
 - i. Meeting time,
 - ii. Agenda
 - iii. How can we do more
 - iv. Should we do more?

Goals

Meeting Minutes

- PA
 - Took data, tried to have same data on both sides
 - of all data, 2 numbers coming out correctly

- different processes and steps to go from stat record to this page
- <link to Peter's doc>
- SDMA table - converts records to table design
- Next action - make sure getting the same # of records starting with all the way thru system
- plan on debugging car years first
- JM - where does data come from?
- PA - raw stat records
- JM - not surprising, goes back to where the difficulty is, discussion couple weeks ago: from requirements or million little things not in the spec, basically repro logic from 4 boxes
- PA - trying to come up with same end result, went over biz layer logic, doing everything at query time, very simple, was built (long time ago), lot of ETL jobs described graphically in ETL tool, cant just read progressive lines of code - lot of decoding, big thing wrap up events in quarters, want to copy that, lets just make it as easy as possible to load w/o mult stages - couple places doing stuff wrong, could be extracting select queries, might have less than = to, maybe greater thans, unpacking and figuring out where deviations are
- DH - consider w/in transmittal table, might want gross prem and incurred losses to see if they come out equal
- PA - woprking with andy to gfigure out various tables, fig out where going differently, expect more clarity
- JM - what I expect, everyone likes stat records, easy to load/und, also that record gets incredibly large, why need steps along the way, idea is that bunch of stuff "we dont want this in code", thinking reverse way back into reasonable number b/w stat records and...
- PA - making report, runnign alias on earned exposure column, lot of code used, start to look at
- KS - as those steps occur, watching how to make it more efficient
- JM - hypothesis - find a lot of repetition, do this in 5 reports will find some quant of repetition - "wouldn't it be handy if we did that once and then replicated" - simplicity can be dangerous concept, if simople means "not as many boxes in pipeline" push complexity to the queries, expect it to be a majority
- PA - constraint = query time doesn't need to be super quick, with one helper module, could recycle the processing module,
- JM - fair tradeoff, which can you debug, pros cons, do all serially in code, no intermediate points - which do have a merit - push it as a layer or set of code modules
- KS - having phys vers of that derived data somewhere after a stage - quartely data, gross prem - whatever they are to test at that level
- JM - performance (not recal on the fly), troubleshooting: if can fix intermediate values vs go at the end
- PA - started off, earned Prem hard to calc, dooing something closer to whats up here, tried to group based on policy and other stuff - other situation we run into is that it would be really nice, make something simple, dont know really 100% sure - both sys running test and data set we know all the values to to make sure this line is equal to this line - no perfect meter stick
- JM - legacy system changes - if you can get to point of balkancing prior thing, if all #s come out how we want them to come out, step on business logic - arguemnt a couple weeks ago, read docs, take reqs, build reqs get right answer - english is context sensitive, computer code is context free - some level chance to do some things in code not spelled out in Reqs - if we can agree to how to document b-rules, in rigiorous effective way may look at rules and think they suck but impressed by navigation - crisp rules, make em better, also take rules and say "highly reusable components", if we can come up with way to doc bizRules really crisp
- KS - make sure we dont lose the submessage peter putting out there - working against data we have in test environ, customer data, complete, matching to actual cust results, impoortant tyo have test data - is there a way to make good test data where we know the results so we dont have to run a full year/records
- DH - will prove legacy reports not correct
- KS - could
- DH - if balances no worse than what it was
- PA - hate to have "code right but broken to make it match"
- KS - we can run test data through old system , a good testbed we canbuild out will be valuable
- PA - w/ what he has implemented, feels close, look at 4 fields calculating, feels confident to explain and point to biz level stuff we have - done a good job of documenting bProcess and rule - where are there checkpoints and audit logs to do more comparisons
- PA - testnet - get into and make 10 records for ea transaction type and get results, by hand or excel, run thru sys, could be val
- JM - build regression test base to keep stable is awesome - if you synth data know right answer is, cert records have well defined outcomes VS grab 3 months of data from real system
- PA - company went out couple years ago, no info on it, test ran and equal results, confirmed all records in SDMA in data set, no extra sub lines
- DH - that dataset, manually decalculate the variables
- PA - 20k records, product used not super easy, thinks excel will handle 25k rows? maybe use excel and calc stuff in excel, doing autofills all the way down, thinking big dataset, realm of possibility
- JM - synth data - purpose of exercise? arg prior systems mostly right, zero out against old system, inherit 90% accurace AND 10% flaws - synt data based on new rules, watershed ? - which path are we trying to use? diff against prior records or synth against new rules
- PA - black box of legacy ETL stuff, run 20 records and do it by hand same results,
- JM - edge of a big directional decision, could ask the resident person doing work dig thru for who knows how long dif down to zero, inherit legacy accuracy and errors, someone has to engineer that - if you can dif against good known dataset - are we saying so confident in new rules vs old rules OR tweak it
- PA - first thought - hybrid approach, dont think got intel to make all the edge cases, not get right on that, expected behavior is coming out correct vs "hey selecting incorrectly" etc, low hanging fruit not dege case based - if i get 10 records to match, worth assuming diff
- DH - bigger picture - reason we need to build to make sure plumbing works from start to end, at this point does it matter oif the report is correct? we need to prove a report can be generated from the statistical record, see info can flow from Carrier to Regulator
- PA - at debugging phase for report gen, confident will be debugged w/ man-hours, when do we start homeowners with this group? und plumbing and AWG work, WHEN do we want to start homeowners? first long meeting in a while? this group?
- KS - heard DH say "we can clarify success criteria for the POC to have inaccurate #s"
- DH - looking at doing things in parallel, need ot have accurate report, but for "worth moving forward POC" that doesnt require matching reports
- JM - agree - shouldnt hold up tech track for numerical accuracy, data scares more than tech, data hard (fun), concern: that logic peter is up against is brutal, long path, get plumbing working but figure out how to do biz logic
- BB - start with a dataset, simplify, dev plumbing, origin was produce accuracy as good as legacy system, start with
- JB - not accepting that data, but resolving issues with report gen
- PA - going through significant re-eval of network interactions and report deliveries and generation - small part is numerical ops, network ops should not be held up
- JM - good enough to tell team "go with plumbing while we do #s"
- BB - if legacy producing incorrect #s for some reason, how can you do the plumbing against something that is incorrect
- PA - developing the networking, slightly decoupled from the calculations on row by row level
- BB - go back to chart, records go thru process, produce legacy report, current effort to replicate legacy report w/ diff approach - what if legacy has errors? re0plicate errors - want to ID errors not repliacte
- KS - run, loing pole from slowing down other efforts, dont know the old #s are wrong, know there is nuance built up over time, going to be finding might be difficult to repro #s, dont know, dont want to put on crit path to prove plumbing

- DH - one module of plumbing, int in how the plumbing interlinks
- JM - tech stack, what tech talking to each other, AWG meets, needs to run, is this sound enough in terms of tech imolications? if data is wrong can still prove tech works, tech track sep thing, dont let data get in their way, continue to fix #s, want to und how much time we work to fix numbers vs defining home UNLESS we have a dependency on peter and it is better that this team spends time on home
- BB - whats th epoint of todays discussion, what questions to be answered?
- PA - update had gen report from both systems, not big diff in quality, update on plan of attach for debugging, discuss test dataset, little worried so many meetings have been recently short with lighyt tech update, afraid wasting business peoples time
- BB - sounds like give an update is there feedback you want from the group
- KS - getting that, JM saying no to Homeowners lets work on #s, beyond update good people who can do good work, is there work this group can do
- DH - continue until 1 plumbing is together or tech team at what point look at homeowners, didn't take loing for auto, maybe month, if it takes 18 months for tech team to move stuff will be stale
- JB - proving #s, PA using MongoDB and some tools there is a platform issue, choices for flexibility down the road
- PA - some of that, for most part tools is gettin direct feedback and synergy, how EPs get processed
- JM - Mongo has challenges, for JM most effective use of time for this group: old code, doing things, better or worse doign things - use of time askl Dev to look at this, inevitably he can tweak something and it will be right, business logic - will find piece of logic, as dev come back to the team with funky logic, peters number baalance is not the goal, we should use this meeting to review this logic, catch new rules, write off garbage - iterative process of finding issues is 1/2 job, other reason is this group questions/answers why is this rule here
- PA - based on business rules, trying to implement alls rules, not sure if failign to implement or if fund diff things going on between two systems
- DH - how big is the dataset?
- PA - 20k records of Prem
- DH - if you send 20k records he can check them
- PA will check rules around data
- JM - key ? report gen via JS did you attempt to write logic with dale's rules? did you write report via JS using dales rules and ran old data or did you reverse engineer old rule?
- PA - tried to implement dales' rules
- JM - watershed - you either let experts synth known data set OR diff (reconcile, could still be wrong) but go back to upper 1/2 of path, make assumption or do the work, "everything that doesnt match must be garbage OR reconcile differences"change rules, change code, diff lower and lower - learn from old system or
- PA - implemented biz rules, couple places with questions, business team with great answer - little things different - value in legacy sys diverging, right now
- JM - do both? DH doing a 20k know set, see answers, if they balance then beautiful, code matches logic - how diff from legacy system, if this code doesnt balance against old system explain why or change rules - are we gonna be that rigourous? or "no we trust rules" prob bit of both
- PA - would like to put in effort with Andy, take records we have, biz rules we like, find checkpoints (SDMA, loading process), find out where/why stuff wrong
- DH - when goes thru SDMA - stat records both edited or accepted or accepted coming off SDMA?
- PA - get thru middle table, follow up and make sure - believes its fgoing to be all the reocrds, keep the load at stage to get load under 5% error then load all records into HDS equiv
- DH - SDAM is post-editing? basically stat records are post edit, starting w/ true dataset
- PA - getting in there, checks, etc.
- JM - SDMA is proprietary? would love to look at it, esp filtering logic, comes down to those if/then statements (critical), logic has a signature, can we dump and look at it, fears where it is screwy
- PA - could be a lot of reasons for it, lot of unknowns, feel strong operating bottom line but top as worked in spring to approve loads from carriers and gen final reports, has no exp with middle section
- JM - come down to logic statements, how severe, not going to type 1 or type 2 exhaustively, no matter what say "just what Dales spreadsheet is good", more good known records means dont need to dig into old stuff, if we had time and we could reconcile, grow records more confident known good answers from DH's sheet we can trust it even though can't reconcile 10%
- PA - been audited, black box, sig off right now, worth working on
- JM - this is a watershed moment in how we arch the data, reconcile past, clarify future, where we cut the line, business logic / rules critical, clear confidence

Discussion items

Time	Item	Who	Notes

Notes

Action items



