

The Data Decoder Ring Lake, Warehouse, Bucket, Staging A Plain-English Map-20261023_160256UTC-Meeting Recording

October 23, 2026, 4:02PM

22m 3s



Julie Rieken 0:10

Good Friday afternoon. Let me just check in the chat and see how the audio is working for everyone.

Great. I think it's working there. So if it is, I'll go ahead and get started and we'll go ahead and get through today. Today's going to be quick and hopefully excellent. So let's go through. Okay. So today we're here. I'm Julie Rieken. I'm the CEO of A2 Association Analytics.

And for the next 20 or 30 minutes, we're just going to talk about four words. Lake, Warehouse, Bucket, and Staging. And these are words you've almost certainly heard in a meeting. And my goal is that by the end, you can explain them to someone without looking at notes. So let's go ahead and do that.

So...

Um...

Let's go to our next slide here. Great. So the origination point of this topic was that we had been on calls with customers and words like these were flying around. Lake and Staging, silver layer. We'd say them and people would talk about them. and everyone would just keep moving on. And I realized in the middle of it, I'm fairly technical-ish, and I have a decent ability to understand where my team's technical skills are at. And I found myself nodding along as if I kind of knew. And I realized I didn't understand the difference between some of those words. So I needed to look into it because they're pretty common nowadays. So when we would debrief after some of these meetings, I did discover that we all didn't know the same thing about the words. And we had different pictures in different heads. So if we were all confused, I thought maybe it'd be important that that we learn what they are. Most of us are just too polite to say that we don't understand what all the words are. So that's what I had found out was happening. And so therefore today. So these were some of the questions that I didn't end up asking when I was in the room. What is a mart versus a lake?

versus a warehouse? Why do you even need a warehouse? And what does Staging do? And what's an S3 bucket? These were things that were holding me back from understanding some of the conversations that were happening around me. So today, you're going to leave with four different words in plain English and the nicknames people use for them.

and one sentence that explains a lake versus a Warehouse. And it's short enough to say in a meeting so that you'll know what the words mean. And a question that you can ask any vendor or any presenter, or when you're at any kind of a presentation and you hear these words, you'll have a better understanding of what they are. So let's just go ahead and start through with some of these words.

So before we get started, we're going to take an analogy and we're going to follow it all the way through. And we'll talk about that throughout the webinar. It'll make things a lot easier. So I'm going to compare some of these things to items that you might find in a kitchen. So just let's start with the idea of a single tomato.

If you have a single tomato, it can show up in a delivery box. And it's going to then eventually go onto a shelf. And then we're going to put it onto a cutting board, and then eventually we're going to plate it and serve it up. It's the same tomato. It's not a different tomato.

The thing that changes is what was decided about it. Did we slice it? Did we dice it? How much? And did we make a sauce with it? Did we put it on a sandwich? So your data is the same thing. Every single layer of your data, no matter where it is, whether it's an export, raw, Staging, or Warehouse, it's the same member.

Every layer is the same, but what changes is how much we decide about that member. What are the columns called? What does an active member mean? And how does it all fit together? So hold on to the tomato idea. Every word we're about to decode is just a different step on the trip of the tomato.

And it's relevant when it comes to the idea of...

of data. Okay, so let's start with the two words that we hear the most, a data lake and a data warehouse. And the difference is about when do you decide what that data means. So a data lake is like a chest freezer. Everything goes in exactly as it comes home. Unlabeled bags, foiled bricks, and we don't decide anything. We just put it on. all into the freezer. So every time you open the freezer and you're holding a frozen block, you're asking, well, is this chili or is it spaghetti sauce? And every single time you go into that freezer, you have to decide.

A warehouse is like a pantry. Things get labeled and sorted out when you put them

away. So later you just reach in and you know what it is because it's been labeled and organized. So somebody did the thinking up front. So it's a pretty big difference between a lake and a warehouse. Neither one is necessarily bad. A lake can be cheap and flexible. A warehouse is organized and quick.

And lots of times you'll have both, but we're going to walk through why this is important to you. Why do you care if your data is in a lake or if it's in a warehouse? Well...

The whole kitchen before we walk through it has four layers. So the export is going to be a sealed box by the door. Just imagine that Cisco truck pulls up and drops some things off at your kitchen. The raw data is when it's on the shelves, but it's still in the packaging. But maybe you've got tomatoes on one shelf and you've got... peanut butter on another.

The Staging layer is when it's washed, chopped, and in your own labeled containers, and the Warehouse is plated from a written recipe. So we're going to follow this same analogy all the way through, and we're going to call, we're going to use a member now, stop using the tomato. We're going to follow it through, and we're going to call her Dana. And every association has a Dana.

So we're going to talk about Staging. It's A highlighted area. It's a really important one. And sometimes it's a place that data can, that people can skip because it's actually the hardest one. And yet it's, and the reason people skip it is because it's the hardest, but it's also the most critical for you to get good results out of your data.

Okay.

And maybe the premise of why do these words even matter? And it's because we have some new technology. We have AI. And when AI looks at data, it matters what shape that data is in. Is it on a shelf? Has it been plated? Is it in a warehouse? Is it crammed into a freezer? Because every single time AI looks at one of those layers, And anybody's product, it doesn't matter. You need to know how is that data, how is AI looking at that data? Because that is how you get reliable answers. So, okay, let's talk about the first layer. It's just the export. So last night, your AMS exports a file and it lands up in the cloud in some storage.

and it's in a folder by date and nothing ever changes. It's just been shipped exactly as the export. Dana is in that file. She's our member. And she's one row in last night's file and she is sent exactly the way the AMS sent her from the file. So what it is, is it's like Dana.

is sort of like the box that got dropped off behind the kitchen. So she's in there.

That's the first one. That's the first layer. It's just the export. It's not necessarily cleaned up or anything else.

OK, then you unpack the boxes, and this is called raw, so the boxes get unpacked and they get put on shelves.

The file gets unpacked into tables that we can search. Now we can search them, but every column is still named the exact way the AMS vendor named it.

So in the kitchen, the groceries or the tomatoes are on the shelves now, but they're still in the packaging and they're in the supplier's words. So one box might say Skippy peanut butter and another one might say Jif peanut butter.

But they're the same thing, just different words. They're both peanut butter. So why is this important? Because what's on the slide, one of them will say, like an AMS in one association, the member status code could be A, which could mean active.

But a different vendor might say the status equals active, which means that they're both peanut butter, but they're labeled a little bit differently. And that matters to AI because AI has to understand what do you mean by active. So the same people using the same AMS may be using it in different ways.

So nothing's been cleaned up or organized yet. But this is an important layer for your data. It's a safety net because if something turns out to be wrong later, you can go back and just put it on a different shelf.

OK, so that's the raw layer.

Now the Staging layer. This one's pretty important. It's the one I want you to remember.

In Staging, your vendor's words become your words. And your association's rules get applied here. So after this step, it doesn't matter if Dana came from one AMS or a different AMS. She's described the same way. Dana's A

means active. And that user D1 of 14 becomes her chapter, maybe the Ohio Valley. So somebody will decide that once and now it's been written down.

So here's the way I think about it in the kitchen analogy. Most families have a recipe that only one person really knows, and the card says a pinch of salt and cook until it looks right. And when your grandma makes it, it's perfect, but when anybody else tries, it's close, but not exactly the same. And that's because whatever that pinches lives in that person's head, but not on the card. And every association has recipes and definitions like that. Active member means something very specific at your organization. And usually one or two people really know that. And when they're on vacation or when they retire, that recipe goes with them. And at a different

association,

an active member could mean something very different. We've seen associations where no grace period is included and they're still active. We've seen them where active means they are included in grace periods. Some people even allow active beyond the grace and put people into arrears and they're still considered active. So this is a really important moment. You have to be able to shelve

The information in a way that is consistent.

So Staging is the day that somebody sits down and writes down exactly how much a pinch of salt is, or exactly what it means to be an active member, so that every time you count it, it's the same.

And why is this important? Because in a data lake, it hasn't been defined.

So AI or anybody could look inside the data lake, which is remember that freezer when everything's in there, and it could get defined in different ways on different days, which makes your data a little bit less reliable or a lot less reliable. If you've staged that data and created a definition for it,

you're going to get consistent results. And that's super important for your accuracy.

So you want to know that. And that's why those meanings mean something to you.

So that when you're working with your products or with your vendors, you need to know, do you have a staging layer? And is it conformed? Conformed meaning have you written the definition down?

Does it mean the same thing every time? And does your system know what it is? So in your kitchen, it means it's organized, it's washed, it's chopped, and it's in its own labeled container. You've taken that data and you've put it into the right place. Okay. This is why it's important because, again, as we've mentioned, 2 associations define active. You're both right. Both associations are right. It's just that people define them differently. And so you want to make sure that you put it into a format that is readable by both and is the same.

so that you can do comparisons and you can put the special meanings on top of whatever it is. So without this layer, without the Staging layer, every single question you have to re-guess what does your data mean. And that's what AI will do. It will guess every single time without a conformed data model.

It's going to look at that data lake and it's going to make a guess every single time. Is this chili or is it spaghetti sauce? Is it chili or spaghetti sauce? And you don't know which one it's going to be choosing. So.

That is really important for us. So, and by the way, AMS systems, they do ship with

blank fields. So every association fills them in differently. And so you might think one is time and it might actually be oregano, right, in this analogy. But once you know what it is, it's just a really important layer for accuracy. So

There we go. Staging matters. Okay, now the warehouse. The Warehouse is super important. It's the final layer. Everything gets combined into one consistent table.

This is where your marts are and the marts are defined. We're going to put only this thing in this bucket. And dashboards and AI read from here and only here.

In an ideal world, they should be reading from a Warehouse.

Can they read a lake? Yes. But again, that's where the inconsistency comes from. So you want to have a warehouse. A warehouse is really important because it's just like an Amazon warehouse. And you can imagine it's really mechanized and everything's put on shelves and compartmentalized. That's what you want. So dashboards and AI will read from a warehouse.

So if Dana is one of your members, remember her data came in, she's counted once, she's counted the same way, and it doesn't matter if the question comes from a dashboard from AI.

So that's, it's sort of the same like a chain restaurant, right? The burger tastes the same in Ohio and Oregon, not because the cook is a genius, but because there's a written recipe and everybody follows it. And that's what a Warehouse gives you. One shape for every answer. So how many active members do we have becomes one count the same way.

done the same way every time. So the chart shouldn't need to know what AMS you run. So here, I mean the organized, the warehouse is the organized layer of your data. So we'll come back to the other stuff in a minute here. Okay.

You might also hear words like bronze, silver, and gold or medallion. These are nicknames for the same layers we just walked through, and they were popularized by a company called Databricks. So the mapping's a little loose. It's a little bit like a coffee size, tall at one coffee shop is the small somewhere else. But these names that you might hear also, the landing zone, bronze, silver, and gold, those are just names that were popular.

by data bricks.

So if somebody says, what's in the silver layer, it's fair for you to ask, what do you mean by that? Because it might be a little bit of a blend of Ron Staging. So those are just some names that come directly from a company called Databricks. OK.

All right. So this is where the Decoder really earns its keep. Two of these words could

potentially mean two things, depending on who's talking. It's like the word biscuit in the US, it's a fluffy thing you have with gravy, and in the UK, it's a cookie. Same word, completely different thing on the plate. So Staging can mean

The translation step we just covered, but in an old project, it might mean a holding area where raw files wait. And A warehouse can mean the organized layer of your data, or it can mean a product, a database that you buy or rent. So we use a warehouse might just mean someone bought a database. It doesn't tell you whether or not your definitions are organized in it. So that's a good thing to ask.

Does Warehouse mean that my definitions are organized? So it's a really fair question to ask if you see those.

Okay, here's some phrases that you're going to hear and let's put it to work. If somebody says we sync with your AMS, let's translate them into the kitchen. None of these are bad things to say, they're just incomplete. So if somebody says we sync with your AMS, that just means grocery delivery. Great.

But delivery doesn't necessarily mean dinner. It's a bag of groceries. It gets to your data into the bucket on the raw layer. So ask what happens after it arrives. Does it get organized? If they say no data warehouse is needed, that means there's no recipe needed, which could mean how reliable is it?

How does it taste the same every single time? Where are your definitions written down for what a member really is?

If it's, if they, if you say, if they say our AI reads your data directly, that's the chef cooking straight out of a delivery box. It's still scallions in one box and green onions in the other. Ask which layer it's reading.

And we're built on a leading data platform. That's a commercial grade kitchen, beautiful ovens, but great equipment doesn't tell you whether or not there's a recipe. So you could ask what's organized in it.

Okay.

If you only ask one question, ask this one, where is our definition of an active member written down and can you show me? This is a really hard question a lot of times to know what, how do we define active member? We've seen organizations even amongst themselves needing to learn what that is. That's a really normal space to be. So

But you want to make sure you know that because one of the first webinars I did was why member means six things at your organization, because it could mean different things at different points in time, and it can be very complicated. So you want to

know that. So a good answer would point to a specific place. Here it is, this is how it was decided, and here's when we decide somebody's a member and when they're not.

And that goes into the Staging layer, so that at the Warehouse layer you get the same the same answer every time.

Okay, and again, why does this matter to us? And I think it matters because of today's world with AI. It's only as consistent as the layer that it reads. Any of us that have used it know that there's no such thing as an answer that comes back exactly the same. Well, I mean, a lot of the pros around an answer can change, but the numbers should be pretty consistent. And the reason they should be consistent is Because if you have staging and a warehouse with definitions, it's conformed and organized. You should be getting some really accurate answers. If you put AI on top of a lake or on top of raw data, that could be that could be really different. So if you pointed it or if you point your AI at a raw data source.

It could mean the member could mean one thing in one situation and a different thing, you know, from a different situation. So you'd get two different answers and that's not what we want. So once I finally understood these four words, I understood why we were building our software the way that we did because we believe in the staging and the conformed

the conformed data because it provides all of our customers with better, more reliable answers. And that's what you're going to want out of the systems that you use as well. So a lake is where it lands. In conclusion, a lake is where it lands. And the warehouse is where it actually means something that's been organized.

And the layer in the middle of that, in between lake and Warehouse, is where your association will say, what does this stuff mean? What does it mean? So I hope this has been helpful for you. And that is our data Decoder. It was just a short one today. Just hopefully that these words are helpful to you now in conversations.

And I wish you all a really good weekend. Thanks.

If there's any questions, I'll just stay for another couple seconds, but otherwise, have a great weekend.

Okay, excellent. Thank you all.