

Workbench AI Query Authoring Interviews Findings

1. Background

DataCentral Workbench enables data consumers to perform ad hoc queries on datasets in Redshift clusters using SQL syntax. In 2025, BDT is taking on a goal to “Integrate conversational Gen AI into DataCentral Workbench to accelerate query authoring.”

From 11/21/24 - 11/26/24, 5 user interviews were conducted with DataCentral Workbench users. The goals of the research included: 1) Understand the benefit of AI writing SQL 2) Discover users’ expectations for AI query authoring in Workbench 3) Determine if users prefer conversational output vs. 1 query output.

2. Executive Summary

- Users primarily use Andes datasets in their work
- Knowing which datasets and columns to use is the main challenge when writing SQL
- AI for query authoring would help to save time and use best practices
- There are mixed opinions on the benefit of AI query authoring for non-technical users
- Users expect AI to be closely integrated in Workbench with knowledge of datasets
- Dataset metadata and schema are the most important information when writing queries
- Storing all previous conversations is not necessary
- Users prefer a single output for SQL, rather than a conversational output

Recommendation: AI should be deeply integrated in Andes Workbench and use its knowledge of Andes to help users select which datasets and columns to use. Pursue a single query output UI with a conversational AI on the left and a single SQL output on the right, with the ability for users to see previous versions of the query. Provide messaging to let users know that the query was generated with AI and may be incorrect. Ensure dataset metadata and schema information is readily available to reference. Store recent conversations for a limited time, but give users the ability to save conversations.

User Story	Priority
As a user, I want AI to be integrated in Workbench and provide dataset and column recommendations, so that I do not need to manually discover which datasets and columns to use.	0
As a user, I want a single query output UI, so that I can continue to refine a single query instead of having to parse through multiple outputs.	0
As a user, I want a message that the AI can make mistakes, so I am informed to trust but verify the generated output.	0
As a user, I want to see version history of my query, so that I can go back and see previous prompts and outputs.	1
As a user, I want dataset metadata and schema information, so that I do not need to go to other locations to get the information.	1
As a user, I want to see recent conversations, so I can see my conversation history for the past 1 month.	1
As a user, I want to save conversations, so I can go back to a conversation at a later time.	1

3. Detailed Findings

Users primarily use Andes datasets in their work. When asked what percent of their work is done with Andes datasets vs. non-Andes datasets, users consistently said that the “vast majority is Andes.”

Participant	% Andes Datasets	Quote
1	100%	"100% Andes, and then using further along from that using temp tables, because I don't like reading raw data source" - ICQA Data & Reporting Analyst
2	80%	"Vast majority is Andes - 80% [in Workbench]. In general it ends up being cut down, more like 50/50" - Sr. Prod Customer Insight Mgr
3	90%	"90% Andes most likely" - Sr. Program Manager
4	80-90%	"Today, I think it's mostly Andes. We might be using Spectr a lot more going forward. 80-90% now [Andes]" - Consultant, Industry Analytics

Knowing which datasets and columns to use is the main challenge when writing SQL. Multiple users described the challenge with "knowing how the different tables connect" and "finding the right field you need." It can be difficult to understand "what the column names actually mean and not having an example output." The actually writing SQL part isn't the issue, it's "how do I actually put everything together." General SQL best practices and optimizations can also be a pain point when trying to write efficient queries.

- "I think probably the biggest challenge that I face is understanding what the column names actually mean and not having an example output...It's something that we frequently have issues with - if it says ASIN, we don't know what level ASIN until we run the query which can take a long time...Minor revisions - making a temp table, minor mistakes like not putting the word table in there like CREATE TEMP. If it was able to recognize that before it ran or gave a detailed error message, I think that would really help out." - ICQA Data & Reporting Analyst
- "Knowing how the different tables connect within Amazon and different keys are and how to connect everything, often the trickiest part. Going and running and query in how you want to join or aggregate stuff, knowing how different tables can potentially connect, what format that's in, how you might want to filter so the query will actually run...Not necessarily writing the SQL, [it's the] how do I actually put everything together piece." - Sr. Prod Customer Insight Mgr
- "Technically I'm a SDE now but don't have a background in SQL. It's helpful to refactor a lot of queries right now. How can I speed this up? How can I reduce the quantity of CTEs to make it more efficient?" - Sr. Program Manager
- "The first thing that's top of mind is organization... Previous queries I've written, how do I find what I need there, can I search tables/schemas I've used previously. Problems I've solved in the past, or teammates have solved in the past. Then I can just reference those queries. That's one pain point. The other is finding the right field you need - I might know something I need is in a certain schema, but difficult to track down a specific field or flag, exists in one table but not another. Locating the right fields I need for my queries is a big pain point. Just sort of general SQL best practices and optimizations. Making sure I'm using dist and sort keys properly, I'm doing efficient joins, and things like that." - Consultant, Industry Analytics
- "I wouldn't say it's what's difficult about writing it, it seems a very logical language. I want to be as efficient as possible. I'm aware that I am never going to be, I have no desire to be a, let's say a SQL programmer. I have no desire. I'm nearly 40 years old when it comes to sitting down now and trying to spend, or reinvest my previous 20 years, I can't. Am I going to spend the next 20 honing my skill set on something that's going to become very much a tool embedded within an AI. No, I'm not, but I need it today." - Sr. PM

AI for query authoring would help to save time and use best practices. When asked the benefit of AI for query authoring, users mentioned it would use best practices and "write some better SQL." It would also save time with its familiarity with the datasets to "intelligently search for right schemas, fields, maybe having access to metadata." They said that "Speed is #1" benefit, so that individuals don't need to reach out to data engineering teams and can self-service their query needs. For existing queries, it would help to have "AI optimize it and make it more efficient."

- "I think it would be really beneficial if it took into consideration best practices for optimizing jobs. People try to query way too much information than they actually need...It would be really beneficial if it were modifying the job, that it would explain the optimization so that people can learn and if it's authoring it would be nice if it had an option to explain what the query is doing and why." - ICQA Data & Reporting Analyst
- "If AI did it for me, then I wouldn't have to do it which would be great, would save me some time. Would write some better SQL. Part of it is the boilerplate stuff that would save time - way to more quickly turn in all the variables in the table rather than having to write them all out or copy/paste them all in [for the SELECT statement]...Where it would be

interesting is if it were familiar with the tables. More familiarity so I don't have to do as much investigation. Even knowing how big the tables are or if I should join them, create temp tables. Basics of the query and if it had some more knowledge about the actual tables rather than having to pull up the documentation on all of them, that would help speed up my process too." - Sr. Prod Customer Insight Mgr

- "Speed is #1. Response time. If you're going to reach out to data engineering team or local SQL expert, it's going to take longer than using some kind of chat. Also personally you learn a lot while putting together good prompts, you have to understand what you're asking about to get the best response. Also helpful from a self help self development standpoint" - Sr. Program Manager
- "I don't have to put a bookmark in somebody's calendar or an appointment in somebody's calendar. I can iterate and describe it as I go. So what I first want might not be the actual thing that is able to be provided or what I first want might evolve through my interactions with it. What I'd really love is I'd love my own BIE one on one person, right? Preferably the most educated, most experienced, most knowledgeable one available to me. Sadly, they all have different skill sets, different experience and different breadths of knowledge. The AI tool has the opportunity of taking all that knowledge, all that experience and all being available simultaneously." - Sr. PM
- "Having AI intelligently search for right schemas, fields, maybe having access to metadata from Hoot or Bumblebee. If I search for something "Is a customer an Amazon business customer in a retail table?" The field maybe doesn't say exactly that, but it will know enough to return the proper field that maybe says isAmazonBusiness or something like that. Having AI help with search would be huge. Having AI help with intelligently doing autocomplete...Not just with tables and fields, but also copilot and some other code complete technologies do would be very helpful...Those are 2 use cases to consider - as a user coming to a blank canvas and writing a query from scratch, that's its own use case...how do I find ad attributed sales for a particular brand, and AI helps build a basic skeleton that maybe needs to be adjusted. That's something AI can really help with, the scaffolding. And the other is coming with an AI/human-generated query and have AI optimize it and make it more efficient. There's certain things in Redshift that can cause 10x time improvement on runtime. Having AI help with those scenarios would be pretty huge and save a lot of money for the company." - Consultant, Industry Analytics

There are mixed opinions on the benefit of AI query authoring for non-technical users. Some users are excited about the prospect of AI for non-technical users to "serve themselves faster," because there is a "dollar amount associated with lag in response time" to data engineering teams to fulfill the request. However, other users are "hesitant and skeptical of non-technical users using an AI query method" due to the nuance in queries and lack of clean datasets which may result in "getting the wrong results back."

- "I think the biggest thing is get it to operations teams...Any way we can get these tools into their hands, messaging, trainings, things like that. We can do it, but getting buy in from operations is going to uplevel them. Let themselves serve themselves faster and get a response at the site at the time at the problem. That's our biggest headache right now, developing that self service mentality or model with the operations side of things...Dollar amount associated with lag in response. If it can be self service on the spot, I can put the palette on this truck and get it all out, customer would get packages on time...Get this out to everybody, this is huge, super helpful." - Sr. Program Manager
- "I'm really hesitant and skeptical of non-technical users using an AI query method...I feel like there's so much nuance in the queries we have to pull, our datasets would have to be incredibly clean and curated. That's sort of the prerequisite for having a non-technical user query our data. So many flags and caveats and nuances you'd have to explain and have in there. Assuming you have a clean set of data, you could share basic stuff with non-technical users...I'd be concerned about getting the wrong results back because they didn't have a super specific query or prompt...If there's some hand holding there and guardrails in place, then that's something we might want to roll out, but would be very hesitant to do it otherwise." - Consultant, Industry Analytics

Users expect AI to be closely integrated in Workbench with knowledge of datasets. Users expect AI "to be when I'm actually in Workbench writing a query" and "closely integrated with the actual workflow." This may be via inline suggestions or on the side of the screen. Users should be able to "put in some sort of request...with these constraints...formatted in a certain way" where the AI would "offer some suggestions in terms of which datasets would be the best." One user said it would be

nice to “select to explain what’s going on, so that the user can learn from it” while another user described Amazon SQLBank’s functionality to create DAG diagrams of dependencies.

- *“I would expect it to be when I’m actually in Workbench writing a query, I would expect it to be in that screen rather than external piece. We have those today with Cedric or other chatbot, you can always go there and work with them. I would expect this to be closely integrated with the actual workflow. As I’m typing, potentially inline suggestions, maybe in the actual screen in Workbench. Whether that’s inline or popup there, something that would directly connect. Me being able to ask a question and reference what I’ve written and make updates or add onto it...Biggest thing I would expect and/or reason I would use it is if it were integrated in the interface in terms of both whether it’s inline suggestions or if I’m asking it questions, so if I ask it questions about table columns, format they’re in, what should I use as a join, if there is some kind of chat there. I would also expect it to fill in my open SQL window with what we’ve talked about. Would have to be integrated, or I would go use an external tool rather than something that’s in the Workbench screen.” - Sr. Prod Customer Insight Mgr*
- *“Step 1 the user would enter a prompt “I want to look at data...” User would put in some sort of request, I want to look at this data with these constraints. I want the output formatted in a certain way. I would expect the chatbot would offer some suggestions in terms of which datasets would be the best to do in terms of efficiency, time to query, reliability, which might be the hard one. What’s the most reliable data source based on data quality metrics or information we may have? A lot of ways to get the same data, but may not all be correct That’s the most questions we get right now, where should we get this data...I really like VS Code, been using Amazon Q on the side of your screen. Really helpful to have that integrated. Seeing that information on the right side of Workbench would be really interesting. I really like the functionality where you can select a snippet of code and highlight it and Amazon Q will give an explanation of the code. Step through what it’s doing, then chat with it and ask questions in the same window as you’re working. I think that’s super super helpful. I find myself with 8k tabs open at the end of the day” - Sr. Program Manager*
- *“With the ability to modify existing jobs to optimize using best practices. It would be nice if it had a little documentation with doc strings in there. Then an option to select to explain what’s going on, so that the user can learn from it and create better queries.” - ICQA Data & Reporting Analyst*
- *“There was a new internal product rolled out called Amazon SQLBank...I thought it was very impressive, they did a lot of cool stuff with AI. I would point the Workbench team in that direction - I think they’re solving some interesting problems and being able to use AI to create a DAG of all the dependencies is really clever...That right here is very helpful - teams can take those diagrams and post on their wikis.” - Consultant, Industry Analytics*

Dataset metadata and schema are the most important information when writing queries. When asked the most important information when writing queries, users mentioned that “table metadata is key” and “knowing exactly what the fields are.” Information that tells them if they are using the correct dataset and columns is critical. One user mentioned that “being able to look at prior queries” that they or their teammates have worked on is also important.

- *“Schema links, partitions, that kind of thing. I like to use Hubble to test queries before I put them in ETL...I think whenever I make the jobs, I do a doc string at the top with some information about the purpose of the job, Hoot links, and under each table is a schema link, under that I have partitions with the type so that I know when constructing a job or editing a job, I can make sure I’m doing my joins correctly” - ICQA Data & Reporting Analyst*
- *“...sometimes that is going and better understanding the additional columns in a table that I’m familiar with, understanding their format, the actual data in there. For ones I’m not familiar with, going and looking at documentation in something like Hoot or if the team has a wiki page and seeing if they have example queries or if they explain a little more detail.” - Sr. Prod Customer Insight Mgr*
- *“The table metadata is key. Knowing exactly what the fields are, knowing the dist and sort keys and that I’m using those properly...Coming back to that first point and organization and being able to look at prior queries. Work I’ve done previously or other teammates on my team or other teams who have maybe solved similar problems. Seeing how they’ve tackled it...things like that is huge.” - Consultant, Industry Analytics*
- *“Correct dataset. Data types is also really important, like date/time. Any suggestions on how to better manage date/times, timezones. That’s one of the biggest ones as well, since our network covers all of the U.S. We have to make sure we’re querying at the right time and most up to date information in the right timezone. One of the biggest*

headaches we face right now.” - Sr. Program Manager

Storing all previous conversations is not necessary. Users noted that it would probably be better to allow users to save specific conversations rather than saving everything, because that “would probably save a lot of cost” and “if it’s a mess of former chats, I’m probably not going to go to that very often.” The ability to “sort and filter that intelligently will be a big part of whether that feature gets used.” In addition, sharing with team members would be a useful feature where “looking at only our teams queries would be a win.”

- *“...if the user could select this conversation is important to keep retention of would probably save a lot of cost instead of save across the board. Some users may not need retention for history. Option for user to say this is something I’ve set up and something I’d like to keep vs. all the rest of my messages are just retention of a day or whatever the session is...I couldn’t see a reason for all chat history to be saved for a certain amount of time unless a user is specifically wanting to save this for X amount of time.” - ICQA Data & Reporting Analyst*
- *“Truthfully not as important when I’m doing these types of things. If it’s general questions or whatever non-query stuff, it’s not as important. I’m thinking about what I’m doing today, I don’t necessarily wish I could go back. If I was asking how to optimize a query, I wouldn’t necessarily go back to see what it did. Often a back and forth with troubleshooting and getting something that works. If I get a final query and get something that works...don’t have as much of a need to go back and reference it. I’m not sure if conversation history is as impactful as it might be for general business questions or doc editing.” - Sr. Prod Customer Insight Mgr*
- *“Log that’s searchable would be helpful...Comes down to intelligently searching and being able to put in a couple key phrases and have those results returned. Over the course of a year or more, it’s going to be pretty jam packed full of chats I have to sift through. Being able to sort and filter that intelligently will be a big part of whether that feature gets used. If it’s a mess of former chats, I’m probably not going to go to that very often...The method now is pretty good where you can share by sharing a link...Just looking at only at our teams queries would be a win and you can’t do that right now...Being able to share a link is usually sufficient, that usually works pretty well” - Consultant, Industry Analytics*
- *“Conversation history would be interesting. One thing that’s super helpful now with Cedric is you can create databases or portfolios...idea of personas or data collections. You can have a data collection of your queries. I know this person generally formats tables from these tables, like the data sorted in this way, how can I adjust the response based on this person’s query database or query past history. That would be really interesting. I just like the ability to have 3 or 4 queries, and want to merge them all together or join them in some way. If you could put them in a data collection or ask AI the best way to merge them...That’s super helpful if we could share those data collections or query database. Right now we use some AWS tool, SQL bank or something like that...If there was some way to do that in Workbench so it was seamless, would be super interesting.” - Sr. Program Manager*

Users prefer a single output for SQL, rather than a conversational output. When prompted with screenshots of ChatGPT (conversational) and Claude (single output) for query authoring and asked which UI they prefer, all users preferred the Claude single query output. Conversational output “sometimes only outputs part of it” which can be a pain point, but single output to be effective users “would want to have the ability to see previous versions.”

- *“I like how Option B has the steps of what you need and it gives you the query on the right side. So it gives you the most efficient way to do it.” - ICQA Data & Reporting Analyst*
- *“I like option B better - it’s more convenient for small things like changing XYZ...chat sometimes only outputs part of it. Here when you have everything open that you’re working on. It’ll show where you made edits and what you changed” - Sr. Prod Customer Insight Mgr*
- *“Definitely Claude (option B). That’s my biggest complaint with Bedrock right now. If you had a chat history going, you could kind of scroll up on left hand side and click response and it would update the query on the right hand side and go back if you wanted to.” - Sr. Program Manager*
- *“I like option B, assuming you can see what you previously asked it and what it gave you. Option B is visually cleaner, I would want to have the ability to see previous versions. If that can’t be done, then maybe option A is better” - Consultant, Industry Analytics*

4. Additional Feedback

Users use Workbench to test queries before it goes to ETL. “If it was able to generate the dependencies at the top, that would be really cool. If you had an option to add wildcards that are used in ETL, that would be cool as well. The one thing I would love to see Hubble have is an option for the user to select the run date as the previous day, typically whenever I’m using Hubble...it’s to test before it goes to ETL.” - ICQA Data & Reporting Analyst

5. Next Steps

The next steps are to iterate on the Workbench GenAI mockups based on user feedback.

6. Appendix: Study Details

6.1 METHODOLOGY

From 11/21/24 - 11/26/24, 5 user interviews were conducted with DataCentral Workbench users. The goals of the research included: 1) Understand the benefit of AI writing SQL 2) Discover users’ expectations for AI query authoring in Workbench 3) Determine if users prefer conversational output vs. 1 query output.

Driving research questions included: 1) What’s the benefit of AI writing SQL? 2) How do users expect AI query authoring to work? 3) Do users prefer conversational output or 1 query output?