

Andes Views P0 Interview Findings

1. Background

Andes Views P0 was launched in preview on 3/31/23, providing dataset owners the ability to create a view from a single input dataset with a subset of rows and columns. Collected user feedback includes pre-launch User Advisory Board (UAB) demos and User Acceptance Testing (UAT), while post-launch feedback was received through the Andes Views BDT Tech Talk. The feature has yet to receive any HMD feedback since its launch.

From 5/23/23 - 5/25/23, 1:1 user interviews were conducted with 5 BDT users who have used Andes Views. The research study aimed to 1) Determine feature roadmap for Andes Views 2) Evaluate users' satisfaction with Andes Views 3) Identify Andes Views pain points and areas for improvement 4) Discover new features users would like in Andes Views.

2. Executive Summary

Theme 1: Users discovered Andes Views by being contacted directly by BDT, helping them to restrict data access to meet their privacy use cases. Users mostly discovered Andes Views by being contacted by a BDT PM, while some users discovered Andes Views via wikis, mailing lists, and BDT tech talks. Andes Views helps users solve use cases around data privacy by restricting consumer access to a subset of columns and/or rows.

- **Recommendation:** 1) Improve marketing and discoverability of Andes Views to dataset owners via other methods (e.g. release notes/alerts in DataCentral) 2) Promote how Andes Views can be used to support fine grained access control

Theme 2: Users want support for more complex SQL, enhancements to the query editor, reduced technical restrictions, and better integration with other areas of DataCentral. Support for complex SQL, such as JOIN clauses, aggregations, and rolling window timeframes, would help users meet additional business use cases. The ability to edit the SELECT statement, a UI for writing queries and dynamic filters, and improved SQL formatting would improve flexibility and usability for different roles. Users are restricted by lack of support for multi-regions, nested columns, and flexibility in view versioning settings. Displaying views on the Manage Providers page, showing the WHERE clause on the view's Schema Attributes, and showing an indicator on the Data Completeness graph when there's an intentional limited timeframe will help to better integrate views with the rest of DataCentral.

- **Recommendation:** 1) Add support for complex SQL like JOIN clauses and rolling window timeframes 2) Allow users to edit the SELECT statement, provide a UI to write queries, and improve SQL formatting 3) Support additional regions besides us-east-1, nested columns, and allow users to choose their view versioning settings 4) Display views on the Manage Providers page, WHERE clause on the view's Schema Attributes, and an indicator on the Data Completeness graph that the limited timeframe is intentional

Theme 3: Overall, users are satisfied with the Andes Views feature for simple use cases, as it saves time by reducing the need to maintain ETLs. Users think that creating a view is "simple and straightforward" and appreciate that views started meeting use cases immediately after it was released. They save dataset owners' time by reducing the need to maintain ETLs, and save data consumers' time by allowing users to consume data without requiring an HR privacy review.

- **Recommendation:** 1) Maintain the simplistic nature of Andes Views as new features are added 2) Implement user behavior metrics to more effectively track time on task

Note: Enhancement user stories can be found in [Appendix 6.5](#).

3. Detailed Findings

Participant	How satisfied are you with Andes Views? (1 - very unsatisfied, 5 - very satisfied)	Role	Comments
1	5	Sr. SDE	For our use case, it's enough.
2	4.5	SDE	I think Andes Views are very great. We tend to use them a lot of the time for PII data.
3	2.5	Sr. Data engineer	I was expecting more in terms of dynamic filtering available in the view, allowing some aggregations...It's limited in terms of its capacity.
4	4	BIE	Because there is room for improvement I would say. But in general, I think it is a very good product and save us a lot of time.
5	5	Sr. Data engineer	The purposes I had and use case I had I would say 5, I was very satisfied. I was able to meet and solve the problem. Views have totally met my requirements.
Average	4.2		

THEME 1: USERS DISCOVERED ANDES VIEWS BY BEING CONTACTED DIRECTLY BY BDT, HELPING THEM TO RESTRICT DATA ACCESS TO MEET THEIR PRIVACY USE CASES.

Users mostly learned about Andes Views via BDT PMs. When asked how they learned about Andes Views, 3/5 of interviewees said that they found out by being emailed by a BDT PM. Participants noted they had been “interested in this functionality for some time” to meet access control use cases. Other ways participants found out about Andes Views includes wikis, mailing lists, and BDT tech talks.

- *“From one of your PMs. We were interested in this functionality for some time and someone told us that it's in development.” - Sr. SDE*
- *“Initially a teammate of mine spearheaded the migration to Andes for our tables. We were also looking at some data access control. Through that I did some research on BDT wiki page. Was able to find information about tagging and Andes Views. So via wiki.” - SDE*
- *“Saw it in an email, engaged by PM...saw views were coming...looking through a wiki and tried it.” - Sr. Data engineer*
- *“I follow your BDT tech talks which happened a few weeks back. I learned then. I also subscribe to the mailing list where you publish announcements and everything. I keep following what you're launching to see if it applies to my organization and if I can leverage data solutions to better serve the pricing stakeholders. Mailing list and BDT tech talks.” - BIE*
- *“I believe [PM] sent out a notice that this was being beta'd. It met a use case that I currently had at that time. So I signed up for the beta.” - Sr. Data engineer*

Users primarily use Andes Views to restrict access to their datasets. Dataset owners create views to abide by privacy regulations and ensure data consumers only have access to data they're allowed to access for their business use case. Data consumers reach out requesting access to the dataset, then dataset owners create a a separate view from the source dataset with the removed columns and/or rows for the data consumer to request access and subscribe.

- *“The most basic use case is we have a large dataset containing information from all services in Amazon from all AWS accounts. Businesses want this information to automate on their side. We can't share the entire dataset, so we create views to filter only their data so we can share only their data to their org.” - Sr. SDE*
- *“Primarily controlling access to specific columns. A bunch of our tables do contain PII data and columns. With Andes Views we can remove that column completely and keep all the other columns and provide that to the customer. Which is great, no need for HR privacy review.” - SDE*
- *“One is privacy concerns. Do not expose our data to everybody and provide limited information. Without maintaining another pipeline, granting access, etc...Cater the data to everybody at the same time but give us a way to control who has access to what.” - Sr. Data engineer*
- *“We are obliged to not share seller non-public information to retail organization. Since we share common pricing metrics...I create a view to scope the data and make it compliant with the retail organization for them to use.” - BIE*

- *"I utilize views to reduce the dataset which was a business requirement by one of our consuming teams, partner teams. To be able to reduce it down to only provide the view or the data that they were permissible to see... That's kind of what I'm looking at views for, really if I don't want an attribute viewed, to remove confusion or privacy or whatever, I just remove it and think of it as a view. So it really is removed from any external consumption outside the team."* - Sr. Data engineer

THEME 2: USERS WANT SUPPORT FOR MORE COMPLEX SQL, ENHANCEMENTS TO THE QUERY EDITOR, REDUCED TECHNICAL RESTRICTIONS, AND BETTER INTEGRATION WITH OTHER AREAS OF DATACENTRAL.

Users would like support for more complex SQL, such as joining multiple source datasets. Some users were expecting more complex SQL and aren't yet using Andes Views because they don't support more dynamic filtering and aggregations. Users have multiple use cases for JOINS, which would help users save time creating and maintaining ETLs. One user noted it would be nice if the SQL supported rolling windows in the timeframe, so the timeframe is not always static.

- *"I was expecting more in terms of dynamic filtering available in the views. Allowing some aggregations...It's limited in terms of its capacity...To be aligned with legal, we have to limit to our datasets to 30 days. We want data longer than that...If we have the capability, we can join with another table...and expose that data as a join. Which is always going to eliminate those customers part of that table. That way, we are not exposing the data for those customers, but we are not limiting the users from 30 days only."* - Sr. Data engineer
- *"Wonderful if I can do some complex SQL or more complex SQL than now. For example, JOINS...For example, I have tables, pricing tables don't have brand information. You grab brand from [another table], I need to join these two tables...For example, for pricing it's very very huge, so joining these tables to get the brand is a time consuming process. If I run weekly or daily and create a table, but this is not scalable...I need to modify the ETL and re-publish the data in Andes and run a backfill. I need to add one field now, and it will take another week to add another field. To make the data available to stakeholders, need to add a field, need to backfill the data...takes time, cost, etc...If I have JOINS, I can include in view without needing to run an ETL."* - BIE
- *"There's multiple datasets, right now I would do that JOIN on the backend and there could be cases here where I want to bring that JOIN in. The problem there would be performance of that JOIN. Where is that SELECT then being executed and the time to produce that data."* - Sr. Data engineer
- *"Other thing I would say is there capability to introduce the rolling window in the view definition. I can filter on a particular column right now, but let's say there's a timestamp column. If I want to limit the data to the last rolling window, 5 days, 7 days, etc. I don't have the capability right now. It should allow a rolling window."* - Sr. Data engineer

Users would like enhancements to the SQL query editor. Users described various ways they would like to be able to create and edit the SQL query when creating a view. One user wishes they could edit the SELECT statement of the view, as it would have prevented them from having to use the API. Other users mentioned a UI for the query so users that don't know SQL can create a view, while another user wanted a UI for creating dynamic query. The query editor itself could have improved formatting to prevent users from copy/pasting queries from third party tools.

- *"There is one instance where we encountered an error while creating an Andes Views. I think it was "invalid view expression...I spoke with BDT team, and one of the columns had to be enclosed. Had to create that view via Swagger...Would be nice to edit the select statement. Not much we could do, running into that error would be much easier to edit here than go through Swagger UI."* - SDE
- *"Probably for someone who doesn't know SQL, maybe can offer more higher level selection. Like don't have them write SQL if they don't need to. If I want to filter by a certain column, allow me to do it in a guided manner. Can have something like create filter on original dataset and instead of writing this syntax here, which in this case is pretty simple, for someone who knows SQL. Someone who doesn't know SQL can probably be helped by a guided process."* - Sr. SDE
- *"We provide them [consumers] definitions, how to do metric calculation on this data. The definition keeps changing, become difficult for people to keep track of definitions. Not everyone is good with implementing dimensions. We have*

such big downstream consumers, we want to help them out. We produce the data at the most granular level, but control metric definition at our end and expose that dataset in terms of dimension and metric. Those metrics and dimension will be collected on the fly will be generated on the fly.” - Sr. Data engineer

- *“This [query editor] is kind of a blob. It’s more of a formatting thing. I actually like this for the fact I generated the SQL outside of the editor and then pasted the pieces into the editor based upon what I wrote.” - Sr. Data engineer*

Users are limited by various technical restrictions of Andes Views. Some users support data across multiple regions, but Andes Views only supports us-east-1 so users cannot fully utilize views. Andes Views also does not support nested datasets, so users who want access to PII data in nested columns cannot be accommodated. Providing the dataset owner the ability to customize how they want their Andes View to be updated based on the latest active version of the source datasets would be preferred to save time manually updating each view.

- *“One of them [pain points] is the fact that it’s us-east-1 only. We have tons of data in other regions. Multi-region support would be amazing, but I know that’s something that wouldn’t be implemented in near future.” - SDE*
- *“There have been instances where a customer reached out requesting access to a table but we couldn’t create the view because the PII was in a nested column that they also needed access to. I would like to have nested column support with this...In those cases we just deny them [requester]. Most of the time they do need the data and we have to say oh well.” - SDE*
- *“If I create a new version of Andes table, I’m not sure if the view will take the data based on the new active version of the table. I suppose if I update it correctly, that I should create another version of the view to be connected to the new active version of the source table. It would be good if actually we can choose a way if the view should connect automatically to the new version of the Andes table. Like with subscription, can decide to connect to new active version or stick to current version. Would be good with views if we did something like that.” - BIE*

There are opportunities for better integration of Andes Views with other areas of DataCentral. One user described multiple ways Andes Views could be better integrated with DataCentral. Users should be able to see which datasets are views without having to go to the table versions details page. The “Visual” tab of the view Schema Attributes does not currently show the WHERE statement, and it would be nice if the Data Completeness graph could indicate the graph is purposefully only populated for a specific timeframe to prevent user confusion.

- *“What are any views I built off of it? It’s under Relationships...How do I see that from a larger provider view?...have to click into the table and see it under Relationships. Say I went to provider, I wouldn’t know what’s a table and what’s a view.” - Sr. Data engineer*
- *“When I look at the view [schema attributes], I have to go to visual, then code. When I look at the visual, what would be great here is if I could see which attribute I filtered on. If there’s on here under description or constraint, “filtered on this value” [Visual tab only shows columns, not where clause]” - Sr. Data engineer*
- *“One use case I have, when I was thinking about building a view that gives a filtered snapshot for it would be filtered for a given period of time. They may only need data for Q1 of this year. In that case, I can see the data completeness of it, but it should say “this is filtered and it’s filtered for this time frame”...It reduces confusion...user says it hasn’t been completed, I can say it’s based on a time based feature only for this period of time as opposed to this past year” - Sr. Data engineer*

THEME 3: OVERALL, USERS ARE SATISFIED WITH THE ANDES VIEWS FEATURE FOR SIMPLE USE CASES, AS IT SAVES TIME BY REDUCING THE NEED TO MAINTAIN ETLS.

Overall, users like the Andes Views feature. For those with simple use cases, the p0 release of Andes Views meets their needs and they were able to immediately start using it to improve their workflows. Andes views reduces users’ time to release and they enjoy the “very simple and straightforward” UI where they can easily see all the columns for a table.

- *“I will give you kudos - views right out of the door it worked and it solved a use case immediately. It actually*

provided me some great value and reduced my time to release. I really appreciate this feature and functionality.” - Sr. Data engineer

- *“We actually started implementing something similar on our own before Andes Views was launched. But now we’re replacing our own solution with Andes Views...For our use case, which is pretty basic, Andes Views is sufficient. We don’t have any pain points.” - Sr. SDE*
- *“I think Andes Views are very great. We tend to use them a lot of the time for PII data.” - SDE*
- *“How you define a view is very simple and straightforward. I don’t miss anything special for this UI. How this is designed is very simple and straightforward. That’s perfect.” - BIE*
- *“What I do like is that I can see all the columns for a given table and I know that here are the columns I need to partition or use as part of a filter.” - Sr. Data engineer*

Andes Views saves users’ time by reducing the need to maintain additional ETLs. Users mentioned it only takes minutes to create an Andes View, saving time for both the dataset owner and data consumer. Dataset owners save many hours every week by not having to create and maintain separate ETL pipelines. Data consumers save time by not having to submit HR privacy reviews to get access to the data they need.

- *“In general, I think it is a very good product and save us a lot of time. Otherwise I would need to create another pipeline, another table...[Andes Views saves] up to 6h for the initial table set up, plus timing to backfill different versions (16h/backfill) and run weekly jobs (2h/week). I cannot think on a valid alternative to views honestly with what exists right now which must be straightforward and simple to be understood by the majority of BIEs which doesn’t necessarily have DE skills.” - BIE*
- *“What I would have been required to do without views is build a completely separate Andes dataset. I would have had to on the load or transform backend be able to filter out just to their requirements and provide that single dataset. With the view though, it was fairly simple, I was able to take this down and say provide me all the attributes and filter down...I would say by doing this it has probably reduced...taken it down from 40 hours to less than 8. Almost a week and making it a day’s worth of task efforts.” - Sr. Data engineer*
- *“It’s a simple column removal, I’m talking less than 3-5 min to get the view set up and send it to the customer...Andes views do not save us time, but save the customer time. I believe that an HR privacy review takes a few weeks to complete, so being able to create an Andes view in minutes greatly reduces customer access time.” - SDE*

4. Additional Feedback

Users want to be able to create views via API. To reduce manual effort, one user said they would like to be able to create views via API: *“We definitely want to restrict access to some of these tables, but we don’t want to have to go in and manually create views every time. If a producer says this data contains PII data, we would love to automatically create a view using an API instead of having to manually go into every table...We don’t need it today, but it’s something in the future we’d definitely like to have so we can have these views created easily.” - SDE*

Users are interested in being able to create hidden source datasets and only expose the views to data consumers. One user mentioned they want to hide their source dataset and only expose views for consumption, so they only need to maintain a single data source: *“I want something to be in Andes, I need to show it in a provider. I want to abstract it or hide it for external views. With views now, I could be under a given provider, presenting the views to our partners or consumers and be able on the backend to hide the actual dataset that is being used to produce those views. If you built that, I would be so pleased.” - Sr. Data engineer*

If data consumers are allowed to create views, then dataset owners need to provide approval. When asked if data consumers should be allowed to create views, one user said, *“They [data consumers] can create it [view], but we should review it. The information here is critical if it’s accessed in its entirety.” - Sr. SDE*

5. Next Steps

The next steps are to prioritize the proposed enhancements with the product team to create an Andes Views feature roadmap based on the user feedback.

6. Appendix: Study Details

6.1 METHODOLOGY

From 5/23/23 - 5/25/23, 1:1 user interviews were conducted with 5 BDT users who have used Andes Views. Participants were recruited based on beta testers of the Andes Views P0 (user list provided by PM). The research study aimed to 1) Determine feature roadmap for Andes Views 2) Evaluate users' satisfaction with Andes Views 3) Identify Andes Views pain points and areas for improvement 4) Discover new features users would like in Andes Views.

6.2 ENHANCEMENT USER STORIES

- UX SEV1 = Trust buster - feature or journey broken
- UX SEV2 = Frustrating - time consuming, overly complex, confusing
- UX SEV3 = Irritant - UI paper cuts, inconsistency, fit-and-finish

UX Severity	UX Issue Type	User Story
1	New feature	As a user, I want support for JOIN clauses, so I can join one or more input datasets to meet my business use cases.
1	New feature	As a user, I want multi-region support for views, so I can create views for datasets not in us-east-1.
1	UI enhancement	As a user, I want support for nested columns, so I can limit user access for nested column data.
2	New feature	As a user, I want support for complex SQL such as computations, so I can create more complex views.
2	New feature	As a user, I want support for rolling window timeframes, so I am not limited by static date ranges in views.
2	New feature	As a user, I want to be able to edit the SELECT statement, so any issues in column selection can be handled via the UI without having to use the API.
2	New feature	As a user, I want a UI to write queries, so non-technical users can easily create views.
2	New feature	As a user, I want to be able to choose my view versioning settings, so I do not have to manually create a new view version when my table has a new active version.
3	New feature	As a user, I want improved marketing for Andes Views, so users are more aware of the functionality Andes Views offers.
3	UI enhancement	As a user, I want improved SQL formatting, so the SQL query is easier to read.
3	UI enhancement	As a user, I want to see with datasets are views on the Manage Providers page, so I do not have to navigate to the table version details page to see associated views.
3	UI enhancement	As a user, I want to see there WHERE clause on the "Visual" tab of the view's Schema Attributes, so I do not have to navigate to the "Code" tab to view it.
3	UI enhancement	As a user, I want to see an indicator on the Data Completeness graph that a view is intentionally for a specific timeframe, so consumers do not think the view is incomplete.