

Andes Views GA Survey Findings

1. Background

Andes Views GA was launched on 8/17/23, providing dataset owners (DSOs) the ability to create a view from a single input dataset with a subset of rows and columns. User interviews for feedback on the P0 release were conducted in Q2 2023 and found that 1) Users discovered Andes Views by being contacted directly by BDT, helping them to restrict data access to meet their privacy use cases. 2) Users want support for more complex SQL, enhancements to the query editor, reduced technical restrictions, and better integration with other areas of DataCentral. 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.

From 8/30/23 - 9/15/23, a survey was conducted with users who have created an Andes View since its launch to 1) Evaluate users' satisfaction with Andes Views 2) Understand use cases for Andes Views 3) Identify benefits of Andes Views 4) Discover enhancements users would like in Andes Views.

2. Executive Summary

The majority of participants are satisfied with Andes Views. Users averaged a 3.58/5 satisfaction with Andes Views. 12 participants said they have "Neutral" satisfaction with Andes views, 13 participants said they are "Satisfied," and one participant said they are "Very satisfied." No participants said they are "Unsatisfied" or "Very unsatisfied." 35% of participants use Andes Views "Daily" and 31% use it "Multiple times a week." Users primarily use Andes Views for fine-grained access control "to provide access for other teams to access subsets of our data that are relevant for their usecase." Views are benefiting users by saving time to create new datasets and maintain data pipelines, reducing costs with "petabytes data saving" and limiting "space reserved on the cluster and data scanned while executing the query," and helping users comply with data privacy requirements. One user said that Andes Views "allows us to improve our overall stance in terms of privacy. Also saves us time because we no longer need to create new datasets for every request. Views are a great way to reuse datasets."

Users want support for JOINS across multiple Andes datasets. Today, views can only be used to filter a subset of data on a single source dataset. Support for joining multiple datasets to create a view is one of the most requested features by users. One user noted that this functionality could save their organization millions of dollars: "This is very important feature for Amazon Private Org which can save millions dollars, we maintain a huge data lake (10+PB) just to ingest everything and filter out 3p data to make data compliant in the APB org and then onboard back to Andes for APB consumers, this is very expensive in storage and ops." Another user mentioned that join functionality would require "making sure the views move automatically with latest base table versions is critical here. Otherwise, managing them would be a nightmare."

Users want consumer-owned views. Views can only be created by DSOs on datasets they own. A top-requested feature is for consumer-owned views, rather than the burden being on the DSO: "This would smooth out back and forths we had to have. If in a 3.0 Access Request, Users could propose what datasets they need, DSO's would merely need to manage/maintain/approve/deny these views but wouldn't need to write the code." Another user echoed this sentiment and said, "I asked one data provider to create the view for us, but he refused to do so. The main reason for rejection was the unwillingness to create the precedent for other data set subscribers. I got the response: 'We have 400+ table subscribers and if everyone ask for the view we will end up with hundreds of views'." Consumer-created views means consumers could "define our own views without having to ingest the entire table to filter using Cradle then onboard back to Andes, we have 200+ tables like this." It would also enable consumers to limit the data they want for their own retention use cases: "BDT is publishing the view on large booker table with retention of 62 days but our requirement is to select only last 15 days data. Currently we are unable to limit the data exposed to the queries on the cluster. as a consequence, if a user tries to select from the large view, then it may waste cluster resources (space, CPU and memory)."

Users want streamlined permissions. Currently, permissions for views are mutually exclusive from the source dataset, resulting in additional effort for DSOs to review access requests to the view. Users noted that "If the customer has permission

on the table, he/she should have permission on the views by default.” They want “automated access grants” and “to be able to bulk upload a list or at least copy over perms from the source table...(would be nice if this option to copy over all permissions showed up at creation).” Another user mentioned there should be additional security for views, similar to data scopes, where DSOs can “push any View Access Requests to our Security team for review. We have this path in Andes 3.0 with data scopes. Need that ability in Views. Views need almost more CR/Security since they’re so simple to create.”

Users want automated version upgrades and inline updates. Today, when the source dataset is upgraded to a new version, the view must be manually upgraded to the new version. Users want automated version upgrades from source datasets, where views should “auto-update to new versions of base table. We update base tables often, and every time we do this, it is really painful to update all views to point to the new base table. Though the views work when pointing to old version of base table, there is a giant red banner on Hoot UI asking our customers to push producers to update the views. When a base table is updated, ideally, I would like 1) A new version of the view with the same permissions+subscriptions as the prior view version 2) The new view to point to the new base table.” Users also mentioned wanting improvements to inline edits, because “Every time the field is added a new version is created, in-line changes should be done in same version” and “it seems that if it’s an inline schema change for source table it should be inline for VIEW.” Another user noted wanting “Better documentation, especially around how subscribers are being notified. Can we ticket them when a VIEW is created or materially changed?”

Users want UI improvements. Users noted UI improvements that could be made to the Andes Views creation UI, including better column selection for large source datasets: “the View UI is not great when I have 300 columns in the source table, and I want to vend 167 columns. Checking and unchecking column names becomes a big problem. I suggest having a free form section or allow a bulk upload of column list feature in the UI for simplicity.” A user also “noticed I was not able to copy/paste in to the UI which forced me to type out my list of Merchant Customer Id’s.” Users mentioned improvements to how Andes Views are displayed on the provider details page, including: 1) show nested views under source datasets: “At the moment when I look at a given provider I just see a whole list of tables and views and it isn’t easy to see which is which, or which view is derived from which table” 2) indicate which datasets are tables vs. views 3) flag unused views. On the view version details page, users want to clearly see information like the 1) supporting code 2) source datasets: “Data sources are hidden in relationships tab these should 1000% be on the overview and crystal clear...this valuable link is really important to connect the dots between sources and the Completeness Graph at a glance.” 3) who created the view.

Users want improved query functionality. The current PartiQL query functionality is missing some functionality and users have noted that “SQL language on where clause can be improved. Row level filtering on views can be optimised.” For example, views cannot be created on datasets with nested columns: “Views with nested columns like array[struct] cannot be queried. This is an Andes3 (dataset & view) limitation, and needs to be fixed. Nested datatypes are normal and have great Redshift support through SUPER datatype. Views need to support them too.” Users would also like to “extract data from JSON objects/arrays in one column and expose them as new columns in the view” and include a “tag to identify primary key.” Another user mentioned that “Revision support and Exclude column support in the PartiQL statement would be a great enhancement and benefit our workflow & customers.”

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

3. Detailed Findings

31 total participants responded to the survey, however 5 noted they do not currently use Andes Views and have been removed from the survey results.

How satisfied are you with Andes Views?	Count (26)	Percent
Very satisfied	1	4%
Satisfied	13	50%
Neutral	12	46%
Unsatisfied	0	0%
Very unsatisfied	0	0%

Participant	How are you using Andes Views? What use cases does it solve?
2	I was trying to find a replacement for Eider to run validation queries on an Andes table table, but it seems like Andes View is more for partial/filtered table subscription than running queries.
3	no data duplication. no need to ingest physical data.
4	Creating column filtered views for compliance reasons.
5	We are evaluating and very excited to onboard with Andes because this feature exists. Use case is being able to have granular access control over a single dataset. We have a dataset that contain data for all our direct consumers, we are separating them by filtering per Andes View, and independently control access.
7	We are using the andes views to create views specific to a groups of people.
9	It allows us to implement row/column level security which is a huge win.
12	We are using them to provide access for other teams to access subsets of our data that are relevant for their usecase. We have a generic workflow dataset containing records for multiple workflow types. The Andes Views allow us to grant consumers access to only the workflow types that are relevant for them.
13	We are planning to evaluate and use Andes views (for Andes 3 subscribed tables) mainly for performance and cost optimization. But until now, BDT has not shared any best practices for optimizing cost and performance using Andes views
14	We are using Andes views to restrict access to certain part of the data in the table (filtering based on a specific column value). This solves purpose of providing restricted access for sensitive dataset and not having physically separated tables.
15	Fine grain access, restricting columns
18	Creating business unit-specific views
19	Reporting for AtoZ application
20	Permissions Team is using it to handle permissions for complex policies like DMA, where we leverage Internal View to enforce such policy.
21	row-level permissions on existing Andes tables
26	We are using Andes Views with reduced retention on Large Booker tables to solve disk space issues and large data scans if the queries are not using right filters.
28	We are using Andes views to share tables where we have to remove a few fields that consumers can no longer have access to.
29	Custom datasets for teams that don't need full data access. We have ~10 views in our Provider. These were built as one-offs for one team. We also are the pilot team for Fine Grained Access Controls. We are using Views to solve a use case where we couldn't filter an internal customer's use case with a data scope without months of BDT work. The Views saved use there. But they need some help, which we'll get in to here.
30	Our team is storing the Content Security Violations reported from our websites in Andes and use Views to query the data for debugging.
31	We are using the views to access less data.

How often do you use Andes Views?	Count (26)	Percent
Daily	9	35%
Multiple times a week	8	31%
Once a week	1	4%
Multiple times a month	1	4%
Once a month	1	4%
Less than once a month	5	19%
Answer not provided	1	4%

Participant	How are Andes Views benefiting you?
3	both time and cost saving
4	Cost / maintenance savings. One extreme example is not having to duplicate an excluded metrics view of Tommy tables representing petabytes data saving.
5	Since Andes 3.0 exists, potentially we could have used the decentralized onboarding, but the freedom of being able to specify where clause as filters from Andes Views and further enhancements, we are betting on this.
7	Maintaining less infrastructure. Time saving and cost savings too
9	Allows us to improve our overall stance in terms of privacy. Also saves us time because we no longer need to create new datasets for every request. Views are a great way to reuse datasets.
12	Enables access restrictions to subsets of our larger table.
13	We are planning to evaluate and use Andes views (for Andes 3 subscribed tables) mainly for performance and cost optimization. But until now, BDT has not shared any best practices for optimizing cost and performance using Andes views
14	Avoiding physical separation of seller's sensitive data.
15	Andes 3.0 access, restricting P&C columns No additional storage
18	This is such a great feature! We provide worldwide reporting tables for a variety of different sustainability initiatives (carbon, waste, water) that all require different access restrictions based on business unit and geography. Having Andes views enables us to easily create these BU-specific partitions without needing to set up individualized Cradle/Datnet jobs. I
19	Integration
20	Andes Views is helping us inch closer to Lagoon views, facilitates our DMA requirement and timeline.
21	Can grant specific permissions to subsets of data without needing a cradle / ETL job to duplicate the table
26	It limits the space reserved on the cluster and data scanned while executing the query.
27	time savings
28	Allows us to quickly restrict the active table with new consumers only being given permissions to the VIEW as we work on migrating existing consumers off the table.
29	It's a decent way to provide data to a one off team. Policies for Fine Grained Access Controls require significant dev time. Views are much more malleable. I was able to create the views in an efficient manner with one exception (#6 below).
30	The cost to store the data in Andes as well as query the data is low. We don't have enough data to determine time savings. Our queries are currently running against non-prod data and are very fast. Am interested to see how long queries run against prod data.
31	Time and cost

Participant	What enhancements would you like to see in Andes Views?
3	1. Andes View can be owned by customer not by the provider. (So we can define our own views without having to ingest the entire table to filter using Cradle then onboard back to Andes, we have 200+ tables like this) 2. Andes view can be joined with other andes tables, this is very important feature for Amazon Private Org which can save millions dollars, we maintain a huge data lake (10+PB) just to ingest everything and filter out 3p data to make data compliant in the APB org and then onboard back to Andes for APB consumers, this is very expensive in storage and ops.
4	1. Ability to create views on tables I don't own. 2. Filtered views based on inner join.
5	Automated access grants, we need this. More granular access control such as access grant to specific columns only would be nice. I know tags feature exists but we can't put custom tags.
7	We would like to create views by joining multiple tables to create a finer grained version of the table. We would also like to create a view over the tables which are not owned by us. Else we would have to create views we will have to copy the data to create such views.
9	Here is my wishlist in order of significance (1) ~~Better version updates~~ Allow views to auto-update to new versions of base table. We update base tables often, and every time we do this, it is really painful to update all views to point to the new base table. Though the views work when pointing to old version of base table, there is a giant red banner on Hoot UI asking our customers to push producers to update the views. When a base table is updated, ideally, I would like - A new version of the view with the same permissions+subscriptions as the prior view version - The new view to point to the new base table (2) ~~Better mechanism to define the view~~ The View UI is not great when I have 300 columns in the source table, and I want to vend 167 columns. Checking and unchecking column names becomes a big problem. I suggest having a free form section or allow a bulk upload of column list feature in the UI for simplicity. (3) ~~Allow merging views~~ You can offer simple inner or left joins on top of multiple views. Starting with 2 views is great, but making sure the views move automatically with latest base table versions is critical here. Otherwise, managing them would be a nightmare. (4) ~~Allow querying nested datatypes from Spectrum~~ Views with nested columns like array[struct] cannot be queried. This is an Andes3 (dataset & view) limitation, and needs to be fixed. Nested datatypes are normal and have great Redshift support through SUPER datatype. Views need to support them too.
12	It would be really beneficial for me if I could extract data from JSON objects/arrays in one column and expose them as new columns in the view.
13	More focus on optimizing cost and performance using Andes views
14	There could be multiple enhancements: - Views should support multiple table joins and filtering - Every time the field is added a new version is created, in-line changes should be done in same version - Views does not show any primary keys but to provide information about granularity to the subscriber, there should be some tag to identify primary key.
15	SQL language on where clause can be improved. Row level filtering on views can be optimised.
17	Have the QS business account compatible with Andes 3.0
18	It would be so great if you all could add a nested view to Hoot that shows how views fit under respective Andes tables on a given provider. At the moment when I look at a given provider I just see a whole list of tables and views and it isn't easy to see which is which, or which view is derived from which table. I think that having all base tables listed by default, with an option to 'Expand to see Views' would be immensely helpful from an organizational standpoint.
19	Better user experience, for searching and finding data
20	Revision support and Exclude column support in the PartiQL statement would be a great enhancement and benefit our workflow & customers.
26	Provide option to subscribers to create or request views to filter or limit the data they want to for their use case. For e.g. BDT is publishing the view on large booker table with retention of 62 days but our requirement is to select only last 15 days data. Currently we are unable to limit the data exposed to the queries on the cluster. as a consequence, if a user tries to select from the large view, then it may waste cluster resources (space, CPU and memory). Permissions on the views can be granted if the customer has already granted permission on the parent table
28	Better management of perms - I would like to be able to bulk upload a list or at least copy over perms from the source table - if they have access to the source table they should have access to the view (would be nice if this option to copy over all permissions showed up at creation)

29	1. The code supporting the view is not displayed clearly on the overview page. Neither are data sources leaving it up to the DSO to use the limited 'description' field for documentation which should be crystal clear to the customer as well as anyone from the DSO that didn't build the View. 2. DSO needs to be able to push any View Access Requests to our Security team for review. We have this path in Andes 3.0 with data scopes. Need that ability in Views. Views need almost more CR/Security since they're so simple to create. 3. Data sources are hidden in relationships tab these should 1000% be on the overview and crystal clear. Stuff this in to the area for 'Scheduled Job Profiles'. For views this valuable link is really important to connect the dots between sources and the Completeness Graph at a glance. 4. The main Provider/DSO 'splash' page (which lists all objects owned by the provider) doesn't give ANY indication that an object is a table, a view, active, deprecated? really anything. That page needs more features than I can describe here, but I digress. Views and Tables should be distinctive in that portal. 5. Managing version upgrades from source tables should probably be improved/automated. Also, in #3 any issues should be flagged on that provider page. 6. Great job on the actual View creation UI, it looks great and is almost there. However, I noticed I was not able to copy/paste in to the UI which forced me to type out my list of Merchant Customer Id's. 7. I'd probably like to see what user created the view somewhere on the page. Just in case we/DSO's have to track that down. 8. Is it possible to flag Unused views from the portal?
30	Assuming Andes Views is Cradle via DataCentral, unless it encompasses all different querying tools. In terms of Cradle, I could see the search searching within the profiles, including the datasets and query itself, being helpful.
31	I would prefer to reduce the retention, as I don't need to access data for more than 15 days in a couple of very large tables. It would be great if you could provide flexibility to customers or offer multiple table versions like _7_days, _15_days, _35_days, _6_months, _2_years.

Participant	Any other thoughts/comments?
3	Feature requests which could save Amazon Private Brand Org million dollars: 1. Andes View support joins with other Andes View. 2. Andes View can be owned by customers not by provider.
6	I asked one data provider to create the view for us, but he refused to do so. The main reason for rejection was the unwillingness to create the precedent for other data set subscribers. I got the response: "We have 400+ table subscribers and if everyone ask for the view we will end up with hundreds of views".
9	Great feature! I would love to see it improve further.
15	Thanks for the initiative on Andes views
28	Better documentation, especially around how subscribers are being notified. Can we ticket them when a VIEW is created or materially changed? Also it seems that if it's an inline schema change for source table it should be inline for VIEW. As a data producer some key functionality seems to be missing.
29	I'm not sure how this would work, but I was thinking about this: consider letting an external team propose a view with DSO control to allow or deny which would pass edit features to DSO only. This would smooth out back and forths we had to have. If in a 3.0 Access Request, Users could propose what datasets they need, DSO's would merely need to manage/maintain/approve/deny these views but wouldn't need to write the code.
31	If the customer has permission on the table, he/she should have permission on the views by default.

4. Additional Feedback

Users want flexibility in data retention. One user noted they want to be able to reduce the data retention, as they “don't need to access data for more than 15 days in a couple of very large tables. It would be great if you could provide flexibility to customers or offer multiple table versions like _7_days, _15_days, _35_days, _6_months, _2_years.”

Users want improvements to DataCentral beyond Andes Views. A few users suggested enhancements not specifically related to Andes Views, such as “have the QS business account compatible with Andes 3.0” and “better user experience, for searching and finding data.” One user noted wanting Cradle search to not only search profile names, but to also search within the profiles: “In terms of Cradle, I could see the search searching within the profiles, including the datasets and query itself, being helpful.”

5. Next Steps

The next steps are to share the findings with the Andes Views team to determine which proposed enhancements should be prioritized for implementation.

6. Appendix: Study Details

6.1 METHODOLOGY

From 8/30/23 - 9/15/23, a FieldSense survey was conducted with users who have created an Andes View since its launch. Users were recruited over email.

1. How satisfied are you with Andes Views?
 - a. Very satisfied
 - b. Satisfied
 - c. Neutral
 - d. Unsatisfied
 - e. Very unsatisfied
2. How are you using Andes Views? What use cases does it solve?
3. How often do you use Andes Views?
 - a. Daily
 - b. Multiple times a week
 - c. Once a week
 - d. Multiple times a month
 - e. Once a month
 - f. Less than once a month
4. How are Andes Views benefiting you? (e.g. time savings, cost savings, etc.)
5. What enhancements would you like to see in Andes Views?
6. Any other thoughts/comments?

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 JOINS, so I can create a view of multiple datasets and do not need to maintain a new data pipeline to create the dataset I want.
1	New feature	As a user, I want consumers to be able to create views, so the onus is not on DSOs to create views on behalf of the consumer.
2	New feature	As a user, I want automated access grants for views, so my access request gets auto-approved if I have access to the source dataset.
2	New feature	As a user, I want to copy permissions from my source dataset to my view, so I do not have to manually approve access requests to my view.
2	New feature	As a user, I want automated version upgrades, so I do not need to manually upgrade my view when the source dataset is upgraded to a new version.
2	New feature	As a user, I want inline changes on views, so inline updates do not require a new view version.
2	UI enhancement	As a user, I want freeform selection or bulk upload for selecting columns, so I do not have to check checkboxes one by one for large source datasets.
2	Bug fix	As a user, I want to copy/paste in the query editor, so I do not need to manually type my entire query.
2	UI enhancement	As a user, I want to see which views are associated with which datasets, so I can easily identify the relationship between my datasets and views.
2	UI enhancement	As a user, I want to see which datasets are tables vs. views on the provider details page, so I know the dataset type without having to view the dataset details page. Note: This is being addressed as part of provider details redesign
2	UI enhancement	As a user, I want to flag unused views on the provider details page, so I know which views are not needed. Note: This is being addressed as part of provider details redesign
2	UI enhancement	As a user, I want to clearly see supporting code on the view version details page, so I know how the view was created.
2	UI enhancement	As a user, I want source datasets to be more visible on the view version details page, so I do not have to go to the Relationships tab to see it.
2	UI enhancement	As a user, I want to see who created the view on the view version details page, so the information is readily available.
1	New feature	As a user, I want to create views on source datasets with nested columns, so I can create views on all types of datasets.
1	New feature	As a user, I want to extract data from JSON objects/arrays as a new column, so I can create views with all the data I need.
1	New feature	As a user, I want revision support and exclude column support, so I can write queries that would benefit my workflows and customer use cases.