Meeting Transcript

Meeting: DataFlow Analytics - Product Roadmap Discussion

Date: December 18, 2024 | 2:30 PM EST

Platform: Microsoft Teams

Participants:

- James Wilson (Customer Success Manager) - james.wilson@ourcompany.com
- Priya Sharma (Product Manager) - priya.sharma@ourcompany.com
- Alex Rivera (Head of Data Engineering) - a.rivera@dataflow.io
- Sophie Martin (Analytics Director) - s.martin@dataflow.io

Transcript

00:00:08 **James Wilson**

Hey Alex, Sophie! Good to see you both again. I brought Priya from our product team today because I know you had some feature requests you wanted to discuss.

00:00:25 **Alex Rivera**

Perfect timing! We've been using the platform for 8 months now and overall it's been great. But as we've scaled, we're hitting some limitations that are becoming real blockers.

00:00:45 **Priya Sharma**

Thanks for sharing that feedback - it's exactly what I need to hear. Can you walk me through what's not working?

00:01:02 **Sophie Martin**

I'll start. The biggest issue for my team is the dashboard load times. When we first deployed, everything was snappy. Now that we're running 50+ dashboards with complex queries, some pages take 30 seconds to load. Our executives are complaining.

00:01:28 **Priya Sharma**

That's concerning. Are you using our caching layer? We recently added query result caching that should help.

00:01:42 **Alex Rivera**

We tried it, but the cache invalidation is too aggressive. Our data updates every 15 minutes, and the cache seems to clear every 5 minutes. We end up recomputing everything constantly.

00:02:05 **Priya Sharma**

Ah, I see the issue. The default TTL is 5 minutes which is too short for your use case. We actually have a feature in development for smart cache invalidation based on data freshness. Let me make a note to prioritize that.

00:02:28 **Sophie Martin**

That would be huge. The other thing that's driving us crazy is the lack of row-level security. We have different teams that should only see their own data, but right now it's all or nothing.

00:02:52   **Priya Sharma**

Row-level security is actually on our Q1 roadmap. We're designing it to work with your existing identity provider - so if someone is in the 'Marketing' group in Okta, they only see marketing data. Is that the model you're thinking of?

00:03:18   **Alex Rivera**

Exactly! We're an Okta shop so that would be perfect. Any chance you could make it attribute-based rather than just group-based? Some of our rules are more complex.

00:03:38   **Priya Sharma**

Can you give me an example of a complex rule?

00:03:45   **Alex Rivera**

Sure. We have regional managers who should see data for their region plus any accounts they personally own, regardless of region. So it's like 'region = X OR owner_id = user_id'. Pure group membership can't express that.

00:04:12   **Priya Sharma**

That's a great use case. We were planning group-based initially, but I'll push for attribute-based policies in the first release. It's more work upfront but clearly needed.

00:04:35   **James Wilson**

Alex, I remember you mentioning some issues with the Python SDK last time. Is that still a pain point?

00:04:48   **Alex Rivera**

Yes! The SDK is honestly pretty frustrating. Error messages are cryptic - when something fails, we just get 'Internal Error' with no context. My engineers waste hours debugging.

00:05:10   **Alex Rivera**

Also, there's no async support. We're running everything through asyncio and we have to use thread pools to wrap your synchronous calls. It's a mess.

00:05:28   **Priya Sharma**

The error messages thing is embarrassing. I'll talk to the API team about that. For async, we have it working internally but haven't released it yet. Would you be willing to beta test an async version?

00:05:52   **Alex Rivera**

Absolutely. We'd be happy to. When could we get that?

00:06:02   **Priya Sharma**

Let me check with engineering. If you're willing to test, I might be able to get you a dev build in 2 weeks.

00:06:18   **Sophie Martin**

Priya, while we're on wish lists - any plans for natural language querying? I have business users who can't write SQL but want to explore data on their own.

00:06:38   **Priya Sharma**

Funny you should ask - we're actually prototyping that right now. Using GPT-4 to translate questions into SQL. Early results are promising but accuracy is around 85% which isn't production-ready yet.

00:07:02   **Sophie Martin**

Even 85% would be useful if there's a way to verify the generated query. Like showing the SQL and letting users confirm before running.

00:07:18 **Priya Sharma**

That's exactly how we're building it. The user asks a question, sees the generated SQL, can edit if needed, then runs. We call it 'assisted querying' rather than 'automatic querying'.

00:07:40 **Alex Rivera**

I love that approach. Some of the fully automated solutions are scary - I don't want an AI accidentally joining the wrong tables and returning incorrect data.

00:07:58 **James Wilson**

How's the migration from your old platform going, by the way? Last I heard you were about halfway through.

00:08:12 **Sophie Martin**

We're actually at about 80% now. The remaining 20% is our most complex pipelines. The biggest blocker is the lack of incremental materialization. We have to either fully refresh or use external orchestration.

00:08:38 **Alex Rivera**

Yeah, we're running Airflow alongside your platform just to handle incremental updates. It would be so much simpler if it was built in.

00:08:55 **Priya Sharma**

Incremental materialization is live in our beta channel actually. We haven't announced it broadly because we're still tuning the merge strategy. But you could opt into the beta if you want.

00:09:18 **Alex Rivera**

Wait, really? That's huge! Yes, please sign us up. Even if it's rough we can work around issues.

00:09:32 **Priya Sharma**

Great, I'll get you added after this call. One caveat - it only works with timestamp-based incremental logic for now. We don't support merge keys yet.

00:09:50 **Alex Rivera**

That covers 90% of our use cases. The merge key stuff is only needed for slowly changing dimensions.

00:10:05 **Sophie Martin**

Here's a random one - any chance you could add support for custom color palettes in dashboards? Our brand team is very particular about colors and the current options don't match our guidelines.

00:10:28 **Priya Sharma**

Ha! You're the third customer this week to ask for that. It's a small thing but clearly important. I'll bump it up the backlog.

00:10:45 **James Wilson**

Sophie, have you looked at the white-labeling options? You can at least customize the logo and header colors currently.

00:10:58 **Sophie Martin**

Yeah, we use that. It's mostly the chart colors that don't match. Our brand uses this specific shade of teal that isn't in your palette.

00:11:15 **Priya Sharma**

Got it. Custom color palettes, incremental materialization, row-level security, async SDK, better error

messages... I'm getting a good list here. Any other burning issues?

00:11:38   **Alex Rivera**

One more big one - data lineage. When something breaks, we have no visibility into what upstream changes might have caused it. We're flying blind.

00:11:58   **Priya Sharma**

Data lineage is in active development. We're tracking column-level lineage now, should be in GA by March. It'll show you exactly which tables and columns feed into each dashboard.

00:12:22   **Alex Rivera**

Column-level is exactly what we need. Table-level isn't granular enough. Will it integrate with external tools? We use DataHub for our broader data catalog.

00:12:42   **Priya Sharma**

We're building OpenLineage support, so any tool that speaks OpenLineage - including DataHub - will be able to ingest our lineage data. It's all about fitting into the modern data stack, not being a walled garden.

00:13:08   **Sophie Martin**

That philosophy is honestly why we chose you over the competition. Some vendors try to be everything and end up doing nothing well.

00:13:25   **James Wilson**

Speaking of the competition, I heard you were looking at adding another tool for ML model monitoring. Any update there?

00:13:42   **Alex Rivera**

We were, but we put it on hold. Honestly, if you added model performance tracking to your platform, we'd use that instead of buying another tool. Less context switching.

00:14:02   **Priya Sharma**

ML monitoring is on our 2025 roadmap. Not immediate, but it's coming. We're partnering with some model serving platforms to make it seamless.

00:14:22   **Alex Rivera**

Even basic stuff like tracking prediction drift would be valuable. We don't need full MLOps - just enough to know when models are degrading.

00:14:40   **Priya Sharma**

Noted. I'll make sure the initial release covers drift detection at minimum.

00:14:52   **Sophie Martin**

Priya, thank you for actually listening to all this. I've been on calls with other vendors where they just nod and nothing ever changes.

00:15:08   **Priya Sharma**

Customer feedback is literally how I prioritize the roadmap. Every feature I mentioned today exists because customers like you asked for it. Keep the requests coming!

00:15:28   **James Wilson**

Alright, let's wrap up. I'll send a summary of today's discussion with timeline estimates for everything we talked about. Does next month work for another check-in?

00:15:48   **Alex Rivera**

Works for us. Hopefully by then we'll have tested the incremental materialization and async SDK.

00:16:02   **Sophie Martin**

Looking forward to it. Thanks everyone!

00:16:10   **Priya Sharma**

Thank you for all the feedback. I'll be in touch about those beta programs.

00:16:20   **James Wilson**

Great call! Talk soon.

End of transcript - Generated by Microsoft Teams

00:15:48   **Alex Rivera**

Works for us. Hopefully by then we'll have tested the incremental materialization and async SDK.

00:16:02   **Sophie Martin**

Looking forward to it. Thanks everyone!