Amundsen Community Meeting Notes

Links:

Amundsen Search - May 2022 Community Meeting

youtu.be

Allison Suarez Miranda (Lyft) presented at our May 2022 Community Meeting and spoke about upda...

12
11
10
9
8
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11

Turn on screen reader support

To enable screen reader support, press Ctrl+Alt+Z To learn about keyboard shortcuts, press Ctrl+slash

Amundsen Search - May 2022 Community Meeting

Zoom link

Slides with agendayy. B h h.

Hackerone - realtess

Future topics:

- Brex (Kimmy) [Jan/Feb/Mar]
- Lendingkart (Srivamsi) [Jan/Feb/Mar]

Reminder for administrator:

- Welcome users
- Manage time
- Ask the audience to raise their hand on Zoom, if they want to ask a question.

2022/05/19

- Updates to Amundsen search (Allison Suarez Miranda, Lyft)
- Nested column type parsing in Amundsen (Kristen Armes, Lyft)

2022/02/03

- Careem Networks (Ali)

- Samsara (Jack Roof)

2022/02/03B h. H.

- Skipped

2022/01/06

- Sandesh (Cimpress Technology)
- Emin Can Oğuz (Hepsiburada)

2021/12/09

- Skipped due to holidays

2021/11/04

- Amundsen at One Medical (Blaine Elliott, Andy Huynh)
- Amundsen at King (Robert Barton and Judy Palimonka)

2021/10/07

- Amundsen at Delivery Hero (Marwan)
- Atlas Integration-AtlasSerializable, OpenLineage Integration, & Pandas Profiling Integration (Dominik Choma & Mariusz Gorski)

2021/09/02

- Skipped due to vacation

2021/08/05

- Adam Boscarino at Devoted
- What's next in Amundsen?
- Dinan at Revolut

2021/06/03

- ML features in Amundsen, Allison Suarez
- Olalekin et al

2021/05/06

Recording: https://www.youtube.com/watch?v=pGbKzzqFHSM

1. Why at Snap (Justin Jordan)

1. dbt integration with Amundsen (Grant & Verdan)
2. Amundsen mono repo (Dorian)
    - 2021/04/01hh h. Hh h h

- Recording: https://youtu.be/IE4T8OBvhEs
- AWS Neptune support (Andrew Ciambrone, SeatGeek) - slides
- Feast + Amundsen integration (Mariusz at GetInData) - slides, blog
- Update on lineage (Verdan at Stemma.ai) - rfc, rfc

2021/03/04hh hh. H h. B. Hh. Hh hh h. H.

Recording: https://youtu.be/7an8uQLY_c0
Attendees: 42

- Previewing BigQuery datasets in Amundsen by Jacob Scherffenberg-Møller(Unity)
- Amundsen at Databricks (Tao)
- What's coming to Amundsen (Roadmap)
    - Lineage (Allison)
    - GCP proxy update (Mariusz)

2021/02/04

- OIDC set up (Nirav Langaliya @ Nisum/ Experian)
- Automating release Hh h. H. H. s: RFC discussion (Allison @ Lyft)

Recording at https://youtu.be/X1dnroiqDg4

2021/01/07 - skippe y. H h y y h. Hh. Hh yh hhb h h yh h. Hy yhh h ybybyy. H hh h hhh. H y h. Y. H. H h b h h. H. H hh h h. Hh h. Hy hbhy. Hy y. H. H. Hh h hb hh. Bh h h h h. H. Hh. Hh hhhh. H. Hh. H h hh. G. H h. H. H h. H. H. Hbh. H. H. H. H. H hhb. H. Hh. Hb. H h. Bh hh hh h y. Bh h. Hh. G. G h h. H. Hh. H. H. H h bhbh h hh. Bh hhh. Hbb h g h h h. Hhh. Hhh. H. d due to holidays

2020/12/03

Recording

- Designs for data lineage (Marcos, Daniel Won from Lyft)
- Discussion: Deploh. yment models for Amundsen hhh. Gh. H h. H h. H h h. Hhhh h. H h h h. H b hhh. H b h hb h. H. G h h. H h. H h. Hh bhh. H h. H h. H hh hh. H. B. H. Hh h. H
  a. Do you customize Amundsen? If so, how do you apply customizations?
  b. Do you manage Amundsen using CI/CD? If so, how and when are deploys triggered?
  c. How do you secure Amundsen: both authorization and firewalling?

2020/11/05

Recording: https://www.youtube.com/watch?v=Pn3Iv_3KeHg

- Brex's journey with Amundsen + Looker integration - Yang & Joe from Brex

REA group's journey with data cataloging - Alex, Abhinay and Stacy from REA group

2020/10/01

Recording: https://www.youtube.com/watch?v=6szFHcjXg84

- Rapido's journey with Amundsen - Palash Das, Rapido
- User learnings hh h. H. H. H. H h. B h. Bhh h data lineage - Daniel Won, Lyft
- Demo of lineage in Atlas - Verdan Mahmood, ING

Notes:
~ 29 attendees
Rapido's journey with Amundsen:

1. Metric definition is complicated are hard to discover

Lineage work:

2020/09/03 h. Hbhbh h. H. Hhh h hh g h. Hh h h

Recording: https://youtu.be/Fy_IRWm_IL0

- Adding resource reports in Amundsen (Mariusz Górski & y. H hh h hh. H b hg. Hh. H. H. H hh h. H. H hhb. H hgg g. G bb. H h h. Hh h h h. Hb. H. G. H h. H h h hh h h. H. Hh h h. H. H. H hy hh h. H. Hh h. H. H. Gh b hb h. Ghh ghDominik Choma, ING)
- Amundsen Frontend overview (Tamika Tannis, Lyft)
- Intuit's journey with Data Discovery (Suresh Raman, Pradeep Sk, Intuit)

2020/08/13

Recording: https://www.youtube.com/watch?v=5PrIx6YObnc

~ 50 attendees

Topics/slides:
Donating Amundsen to LF AI (Lyft)
Tableau integration (Carter Landis, Gusto)
Gremlin Proxy design (Josh Hoskins et al, Square)

Announcements:
New #frontend channel in our Slack

Tableau integration notes:

- Gusto launched internally ~2 months ago

- Common questions are around, can I use existing tables and dashboards. Didn't have any dash <-> table lineage
- We believe that Amundsen is the right place to have lineage between data and data/reports, and become the one stop shop for data.
- No equivalent of what Amundsen calls dashboard execution in Tableau
- Only took 2-3 to extend Amundsen dashboards to integrate with Tableau

2020/07/09

Agenda:
Integrating Amundsen with Redash (Jon Hehir et al at Asana)
User research learnings (Parisa at ING)
Amundsen roadmap update (Daniel and Tao)
h. H h h. Hhbh. H hbhh h. H. H. H. B. Hh h. Hh. H hb hhh. H h. H h hh. Hhh h h. H hb ghb h g. H. H h g. H h hhh. Hh. H hh. Hh
Recording: https://www.youtube.com/watch?v=bJTMW2NDjmI
H. H h. H h h. H. H. H h hh. H. Hh b. B. H hh b h. H
Notes:
~33 people attended
H h. Hb. B. H. G h. H. H. Hh h. H hh h
Asana (Jonathan Hehir, DE, Asana):
- Redash - OS, recently purchased by Databricks

Presentation link:
https://docs.google.com/presentation/d/1tiV3sn4B4cNr0H-FAJ7U8ZQ6Yf3n83GUKOQl2GxPhig/edit?usp=sharing

Questions:
- How (if at all) you ensure no pii/other secure values are present in dashboard queries? For example - if someone created dashboard filtering table by Name, Surname and Address of a particular person
    - Lyft and Asana don't deal with this right now.
- At Lyft and Asana, the functionality of calculating view count is internal.

ING (Parisa Khanipour Roshan, UX Researcher, ING):
- DAP (Data Analytics Platform)

Presentation link:
https://docs.google.com/presentation/d/1AsIi_I_0TW5Qf4SOayv0-mvH7_2jDptH9tn_Ynsvkvw/edit#slide=id.g8bdbc1d1b5_0_22

Questions:
- What are you acting?
    a. Improving sorting and relevancy of results
- Surprising?
    a. Wasn't expecting anyone to use other people's tables

- Has ING considered integrating with Superset to preview and explore?
- Official tables - how are they defined?
    a. They are collected from the data lake.
- What's the main barrier to adopting/using tags?

We are not allowing users to add tags y hh h. Hhh het.

    a. Based on this research, we have identified categories of tags. Whether it has PII, we have reached out to data owners to provide these tags to us.
- How many DS and Analysts are in the audience ING?
    a. 450 users on DAP in total, 100 unique per day.
    b. Most of them are on the technical side. Because the tooling that worked well first was Jupiter lab and IDE.

Daniel Won (PM, Lyft):

Presentation link:
https://docs.google.com/presentation/d/1I1ZaATQdXf8k4wBf-shlix8cPnkojhHNPw6GrgXdRUo/edit?usp=sharing .
Questions:

- We (ING) are working on DQ profiles/reports. We can kick on report name and we can materialize and these reports will be
  Idea is to return the links from the report to H. Hhh. H. H. H the metadata service, so as to navigate
- How would data lineage be constructed?
    - At Lyft, we have a central query log that we will parse.

h. H. Hh. H. H b h

- Roadmap mapping
    a. History of the roadmap page, so that's a part of the plan
        ■ (for now the poor persons version: https://github.com/lyft/amundsen/commits/master/docs/roadmap.md)
- Release management
    a. Release page on the frontend app
    b. https://github.com/lyft/amundsenfrontendlibrary/releases

2020/06/04

Agenda:

- Improvements to search relevance (Damian, Verdan, b h h hh. H. Hh. B hhhh. H. Hh. H h h h. H h h. H , ING)- Slides
- Popmon - DQ system (Max Baak, ING) - Slides, Source code
- CLI based on Amundsen (Robert Yi, Airbnb) - Source code, Slides

1. **Recording:** https://youtu.be/b h. Hhh. H. H h h hh. H. H h. H. Hh. Hh. H. H h h. H. H. Hh h. H hbh h h. Hy. H h h. Hh h bhh. Hh. Hh. H. Hhh. H. Ghh h hh hh. H. Bhh h. H h. Hh. Hh hh. H. Hh hh. H h. H. Hhh h. Hb. H. Hh. H. Hh h g. H h. H h hh h h h. H. H h. H y h. Hh. H hbh. H h. Hh. H h. H h g h. H h h. Hhhh h. H. H. H. Hh. H h. H. Hhh h. H h h y. H h. H. H. H. H. H. Hh h. Hh. Hh h. H. Hhhh. B. H. H. H h h h h g.

H h. Gg h h h h. Hhbh h. H h. H h h hh hh h hhh. Hhh h h hhh. H h. Hh. Hhh. H h. H h. H. H hhh. H M7wMefOfJCM

Attendees list:
~30 peoplehttps://amundsenworkspace.slack.com/archives/CGFBVT23V/p1591298198153500

2020/05/07

Agenda:
- Dashboards in Amundsen demo and other future entities (Lyft)
- iRobot's implementation of Amundsen (Michelle Gulen and Christina Leber/iRobot)
- Integrating Data Quality system with Amundsen via programmatic descriptions (Sam Shuster/edmunds.com)
OIDC demo (Jørn A Hansen/Bang & Olufsen)

Recording: https://youtu.be/gs9KYxMjmGk

Dashboards in Amundsen (Daniel Won, Jin Chang, Lyft)
- Currently only indexing Mode dashboards.
- Can search for dashboards
- Can bookmark dashboards
- Information about the dashboard
- Documentation request - what API is being - Jin Chang to update

Amundsen at iRobot (Michelle Gullen , Christina Leber TL, iRobot)
- iRobot is the creator of Roomba
- Internal data users had a lot of pain points
    - 57% of people said awareness and definition of data sets
    - 65% wanted endorsement or warning
    - 70% wanted documentation
- Thanks to Amundsen makes data discoverable, eliminates the manual effort, it's very collaborative by nature.
- Changes made to Amundsen:
    - Partitioned
    - Table size
    - Record count
    - Date created
    - Last Updated
    - Date Range
    - Preview in Mode
- Launched in Dec 2019
    - Indexed several hundred tables
    - "Wish Amundsen existed when I started at iRobot, this is amazing" - Data Scientist at iRobot.
    - Average ~20 searches a day
    - 600+ edits to owners, descriptions and tags
- Common user feedback
    - Search results are pretty broad, trying to make search results more helpful.
    - Missing descriptions, need a better way to track what's missing.

    - More table metadata,
- Learnings
    - Data Scientists are primary folks adding descriptions
    - Want to get more eng to write descriptions
- How are you getting others to add metadata?
    - Folks that are upstream in the data collection process. They don't use the data so they don't see that it's missing.
    - There is a feature that where users can request a description.
- How do you keep your code base up-to-date with the open source community?
    - Prior conversation from Ashhar Hasan (here)
    - It's definitely a struggle.
    - When we started working, we did a fork of the code base.
    - We are now in process trying to pull in new updates to the open source repo
    - Working on a new process for how to do that on a more frequent basis.
    - One request for the community - when something like a new feature has been released, what's the communication channel that we should use to figure out
        - Current best places:
- How are you calculating usage metrics?
    - Merging the action logger that was already a part of Amundsen.

DQ integration with Amundsen at Edmunds.com (Sam Shuster, Edmunds.com)
Current process for DQ check:

- Metrics
- Metrics values
- Tests
- Reports

Now we have very useful metadata on tables. One of our goals is we'd like Amundsen to be single source of metadata in the org. Perhaps, Amundsen will point to pages to internal tool at some point.

There was a missing piece of the puzzle. This is now known as programmatic descriptions.
A normal description is a special case of a programmatic description.

When to use programmatic descriptions?
When you have an authoritative source of metadata somewhere else, that's not Amundsen.

- Example integration

Using for legal restrictions. We want the legal team edit that data, since Amundsen is not editable.

- Integration with Data Quality
    - There's a traffic light system
- Future needs

OIDC connector with Amundsen (Jorn Hansen at Bang & Olufsen):
Open ID connector for Amundsen.

- User Experience

    - Goal: well known, trusted system to log in.
- Owner Experience
    - Minimal work to read, etc.
- Developer Experience
    - Existing original detailed docs on how pieces fit together. For if you need to customize, extend or maintain.

Announcements/Open discussion:

- Community meetings first thursday of every month.
- Contribute agenda items.
- Share feedback on the community meeting

Attendees (46 people):

Brandon Ruhl9:01 AM
Hi I'm Brandon, I am a data engineer from CBS Interactive.

I am Mark, work on Amundsen at Lyft.

John Yoon9:02 AM
Hi I'm John from Geotab, have been building catalog POC with Amundsen since last month

I'm here to learn more and see how others use the product

Sam Shuster9:02 AM
hey i am sam from edmunds.com

Jørn Hansen9:02 AM
Jørn (or Jorn) from Bang & Olufsen

Ashhar Hasan9:02 AM
Hi, I'm Ashhar, working as a software engineer on the data platform at Delhivery in India.

Daniel Won9:02 AM
Hello, Im Daniel from Lyft. I work on Amundsen

David Cameron9:02 AM
Hi! I'm a Data Science Engineer in Security at DigitalOcean. We are implementing Amundsen to help us communicate our data use policy internally. This is the first time I've joined a community meeting.

Mariusz Górski9:02 AM
I am Mariusz, working from Poland as a Data Engineer in ING WBAA. Actually teammate with Verdan, extending the Atlas integration.

Daniel Mateus Pires9:02 AM
Hey :), Daniel from Earnest Research here, we recently did an internal hackathon on Amundsen, been following the project for a while now

Antonio Scardovi9:03 AM
I'm Antonio, working as a developer on iRobot's implementation of Amundsen

Lorena9:03 AM
I'm Lorena, work in data at a start up. We're looking into data catalogs and data discovery platforms.

Glad to be here and see how people are using Amundsen in their orgs

Adam Boscarino9:03 AM
Hi, I'm Adam. I'm a Data Engineer at Devoted Health. We recently started using Amundsen.

Michelle Gulen9:03 AM

Michelle, manager of data analytics and sponsor/internal PM of Amundsen at iRobot

Gee Yeol Nahm9:03 AM
Hi Gee here from WeTransfer. Started testing Amundsen to integrate it into our system.

You9:03 AM
Notes are being taken here:

https://docs.google.com/document/d/1bsJWNt1GBFmV-aRbHFuYgMFnMgIlvAmbhsvDatb0Vis

Tao Feng9:04 AM
Hi, I'm Tao. I work on Amundsen and Airflow at Lyft.

James Davidheiser9:04 AM
Hi, I'm James - Data Infra at Asana. We just started trialing Amundsen in Kubernetes and are excited to start sharing it out with our stakeholders.

Tamika Tannis9:04 AM
I'm Tamika, working on Amundsen at Lyft. (ttannis on Github)

Knowl Baek9:04 AM
Hi, I'm Knowl, a new Product Designer at Lyft starting to work on Amundsen enhancements and new features. Great to meet you all!

Jin Hyuk Chang9:04 AM
Hi, I am Jin a software engineer from Lyft. Mostly working on Amundsen backend.

Daniel Dicker9:05 AM
Hello! I'm Daniel, a software engineer from Convoy working on implementation of Amundsen internally

Jake Robert Mongaya9:06 AM
Hi I'm Jake, working as a Data Engineer and exploring Amundsen

Ashhar Hasan9:10 AM
Is the dashboard system pluggable? ie. can something like Superset or Redash be used instead of Mode.

2020/01/23

Notes
Video

2019/11/05

Notes

2019/09/05

Notes

2019/07/30

Notes