

Final Presentation

Summer @  amuselabs

Time Series Analytics

by Karnav Popat
under Manu Awasthi

Contents

Problem Statement

Process

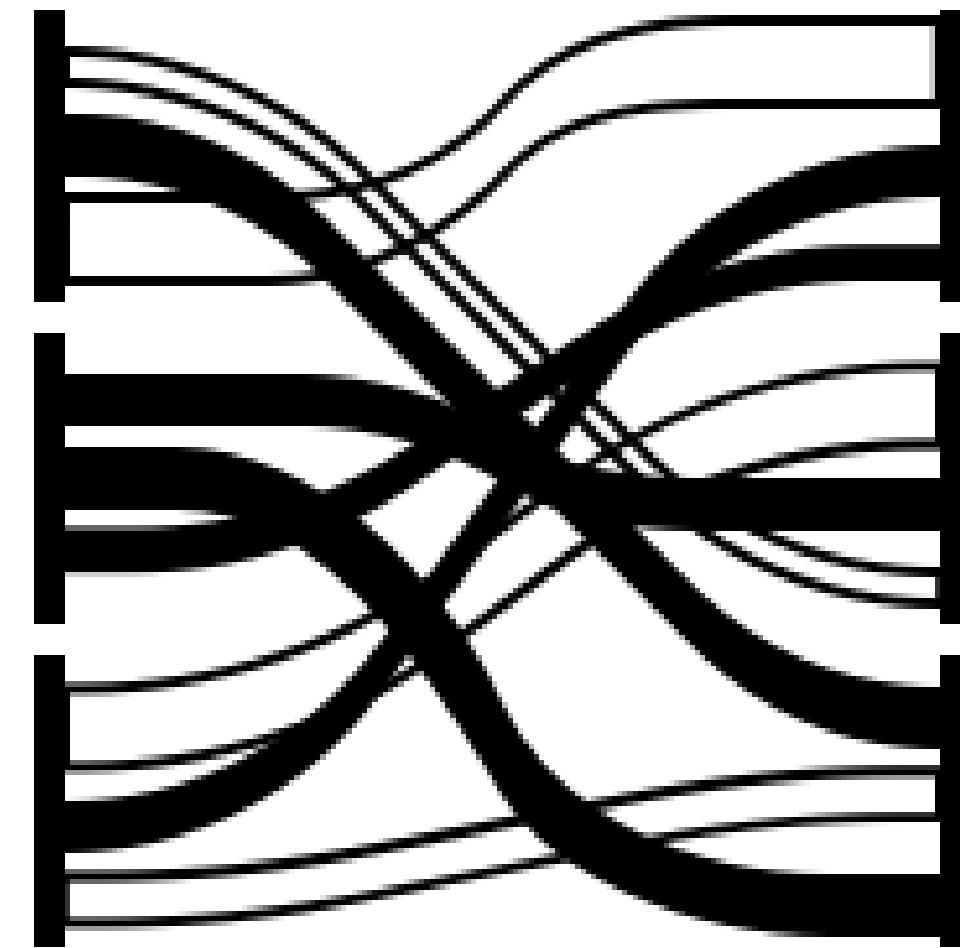
Solution Spec

Learning

Future

Problem Statement

A mountain of data



Problem Statement

There is a large amount of potentially useful information logged every day. This is parsed and condensed into daily, weekly, and monthly email reports, then deleted. Is there a better way of extracting value out of this information?

Any prospective solution must be:

- 1 More useful than the daily & weekly email reports
- 2 Easy to parse, visualize, and draw insights from
- 3 Provide trackability and analysis over time
- 4 Reasonably performant and scalable
- 5 Extensible as the logs and apps change
- 6 Accessible locally without exposure over the internet

Problem Statement

```
1140776 05 Jul 17:48:15.930 DBUtil INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-weekly","id":"69ff9654","playId":"5e898f57-032e-4020-838a-8d72fd64b9e8","userId":"f786c6a3-87d8-42aa-b0af-cf9ab007df67"} with params id: 2cb6f986 set: tny-weekly embed: 1 compact: 1 maxCols: 2 src: http://www.newyorker.com/auth0/postScore
1140777 05 Jul 17:48:15.931 JSPHelper INFO - COMPLETED Request ip: 73.222.124.121, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/crossword referer: https://www.newyorker.com/? with params id: 2cb6f986 set: tny-weekly embed: 1 compact: 1 maxCols: 2 src: http://www.newyorker.com/auth0/postScore
1140778 05 Jul 17:48:16.068 DBUtil INFO - status: success for addOrUpdatePlay playId: 5e898f57-032e-4020-838a-8d72fd64b9e8 uid: f786c6a3-87d8-42aa-b0af-cf9ab007df67 series: tny-weekly id: 69ff9654 rowForPlayMustExist true updated row
1140779 05 Jul 17:48:16.075 DBUtil INFO - setLoadAsVerifiedAndUpdateId uid: f786c6a3-87d8-42aa-b0af-cf9ab007df67 set: tny-weekly timestamp: 1720201695232 nAttempt: 1 UA: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/126.0.0.0 Safari/537.36
1140780 05 Jul 17:48:16.075 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 6.695186 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 6.61563
1140781 05 Jul 17:48:16.075 LoggingFilter INFO - COMPLETED Request ip: 172.58.167.70, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/tny/crossword?id=69ff9654&set=tny-weekly&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140782 05 Jul 17:48:16.118 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"atlantic","id":"atlantic_20240701","playId":"73bfeb53-897e-460b-b038-668216d13cc8","userId":"786c73a6-8b29-4e18-9db7-92e95875f1e"} with params id: 73bfeb53-897e-460b-b038-668216d13cc8 set: atlantic_20240701 rowForPlayMustExist true updated row
1140783 05 Jul 17:48:16.125 DBUtil INFO - status: success for addOrUpdatePlay playId: 73bfeb53-897e-460b-b038-668216d13cc8 uid: 786c73a6-8b29-4e18-9db7-92e95875f1e series: atlantic_20240701 rowForPlayMustExist true updated row
1140784 05 Jul 17:48:16.125 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 6.463916 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140785 05 Jul 17:48:16.125 LoggingFilter INFO - COMPLETED Request ip: 69.138.25.162, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/atlantic/crossword?id=atlantic_20240701&set=atlantic&embed=1&picker=date-picker&theme=atlantic&limit=1
1140786 05 Jul 17:48:16.223 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-weekly","id":"4b4ad6a9","playId":"16b5390f-0268-4f36-b188-f97967a45605","userId":"fa28cb44-e80e-4fdb-9f9f-bde10ee2c3bc"} with params id: 16b5390f-0268-4f36-b188-f97967a45605 set: atlantic_20240701 rowForPlayMustExist true updated row
1140787 05 Jul 17:48:16.230 DBUtil INFO - status: success for addOrUpdatePlay playId: 16b5390f-0268-4f36-b188-f97967a45605 uid: fa28cb44-e80e-4fdb-9f9f-bde10ee2c3bc series: tny-weekly id: 4b4ad6a9 rowForPlayMustExist true updated row
1140788 05 Jul 17:48:16.230 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 6.783844 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140789 05 Jul 17:48:16.231 LoggingFilter INFO - COMPLETED Request ip: 49.199.152.9, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/tny/crossword?id=4b4ad6a9&set=tny-weekly&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140790 05 Jul 17:48:16.409 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"atlantic","id":"atlantic_20240705","playId":"b851cdcf-76ea-4510-8df7-c85198f7202e","userId":"d1a9c3ec-047e-49f6-8f1c-67f9ff799d1"} with params id: b851cdcf-76ea-4510-8df7-c85198f7202e set: atlantic_20240705 rowForPlayMustExist true updated row
1140791 05 Jul 17:48:16.415 DBUtil INFO - status: success for addOrUpdatePlay playId: b851cdcf-76ea-4510-8df7-c85198f7202e uid: d1a9c3ec-047e-49f6-8f1c-67f9ff799d1 series: atlantic_20240705 rowForPlayMustExist true updated row
1140792 05 Jul 17:48:16.415 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 5.9162 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140793 05 Jul 17:48:16.415 LoggingFilter INFO - COMPLETED Request ip: 50.46.233.95, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/? elapsedTime(ms): 6
1140794 05 Jul 17:48:17.804 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":false,"updateLoadTable":false,"series":"tny-weekly","id":"69ff9654","playId":"80869e70f-412a-4dad-9f66-70cdf4c57661","userId":"bfef9d32-f89c-4d9f-b891-3632154b1e76"} with params id: 80869e70f-412a-4dad-9f66-70cdf4c57661 set: atlantic_20240701 rowForPlayMustExist true updated row
1140795 05 Jul 17:48:17.804 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 0.0 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140796 05 Jul 17:48:17.804 LoggingFilter INFO - COMPLETED Request ip: 107.122.189.55, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/? elapsedTime(ms): 0
1140797 05 Jul 17:48:17.805 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-weekly","id":"c6b9cdd5","playId":"15aebb6bc-fe9a-4a1f-9ee2-fadb152dc27","userId":"aaa2b9a2-7a84-4382-8b28-828ed2e7d003"} with params id: 15aebb6bc-fe9a-4a1f-9ee2-fadb152dc27 set: atlantic_20240701 rowForPlayMustExist true updated row
1140798 05 Jul 17:48:17.806 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":false,"updateLoadTable":false,"series":"tny-weekly","id":"00cb1030","playId":"f23e5897-c849-46f7-a8bd-883406cb8851","userId":"0e60dc8c-c97d-45f4-9acb-162f99b71a09"} with params id: f23e5897-c849-46f7-a8bd-883406cb8851 set: atlantic_20240701 rowForPlayMustExist true updated row
1140799 05 Jul 17:48:17.818 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 0.0 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140800 05 Jul 17:48:17.818 LoggingFilter INFO - COMPLETED Request ip: 142.59.68.219, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/? elapsedTime(ms): 12
1140801 05 Jul 17:48:17.817 DBUtil INFO - status: success for addOrUpdatePlay playId: 15aebb6bc-fe9a-4a1f-9ee2-fadb152dc27 uid: aaa2b9a2-7a84-4382-8b28-828ed2e7d003 series: tny-weekly id: c6b9cdd5 rowForPlayMustExist true updated row
1140802 05 Jul 17:48:17.821 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 14.364372 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140803 05 Jul 17:48:17.821 LoggingFilter INFO - COMPLETED Request ip: 208.90.143.199, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/? elapsedTime(ms): 16
1140804 05 Jul 17:48:17.991 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-weekly","id":"0c808f50","playId":"9eb992ae-ef9b-4a57-b9a9-d4c039387973","userId":"72c7a64b-e114-4c5e-9f60-932904743476"} with params id: 9eb992ae-ef9b-4a57-b9a9-d4c039387973 set: atlantic_20240701 rowForPlayMustExist true updated row
1140805 05 Jul 17:48:17.998 DBUtil INFO - status: success for addOrUpdatePlay playId: 9eb992ae-ef9b-4a57-b9a9-d4c039387973 uid: 72c7a64b-e114-4c5e-9f60-932904743476 series: tny-weekly id: 0c808f50 rowForPlayMustExist true updated row
1140806 05 Jul 17:48:17.998 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 7.242102 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140807 05 Jul 17:48:17.999 LoggingFilter INFO - COMPLETED Request ip: 143.58.214.79, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/tny/crossword?id=0c808f50&set=tny-weekly&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140808 05 Jul 17:48:17.116 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-2023-puzzles-issue","id":"84b1dd99","playId":"746481c0-0d60-49de-82ea-51c389feb467","userId":"f83c4d9b-7e4e-4335-9436-37f7d4"} with params id: 746481c0-0d60-49de-82ea-51c389feb467 set: atlantic_20240701 rowForPlayMustExist true updated row
1140809 05 Jul 17:48:18.124 DBUtil INFO - status: success for addOrUpdatePlay playId: 746481c0-0d60-49de-82ea-51c389feb467 uid: f83c4d9b-7e4e-4335-9436-37f7d4 series: tny-2023-puzzles-issue id: 84b1dd99 rowForPlayMustExist true updated row
1140810 05 Jul 17:48:18.124 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 7.640304 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140811 05 Jul 17:48:18.124 LoggingFilter INFO - COMPLETED Request ip: 172.89.76.107, 140.248.67.94, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://www.newyorker.com/projects/interactive/amuse/crossword?id=84b1dd99&set=tny-2023-puzzles-issue&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140812 05 Jul 17:48:18.137 DBUtil INFO - addPickerLoadToDB for addPickerLoadtoDB uid: a54aa6e0-0e1b-45c7-bd37-22b1cf542535 ipAddrHash: ce636ed18494e0f7984a5e833b91532d8db7ff0ce34aa9fc0a0ba22b4ea42945 series: vox id: null updated_row_count: 1 DBtimeTaken(ms): 6 dbName: us_pmdb dbIp: 10.26.64.3
1140813 05 Jul 17:48:18.137 DBUtil INFO - Encoded picker token: eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9eyJiaWQid0iJhNTRhYTZlMC0wZTfLTQ1YzctYmQzNyDyMmIxY2Y1NDI1MzUiLCJ1cGRhdGVMb2FkVGfibGUi0mZhbHNLLCj0aW1lc3RhbXAi0jE3MjAyMDE20TgxMjV9.1qX80gnk0mWJu2BH0vhf92zTyCiDAPX5rt7T3H5ZwQ uid: a54aa6e0-0e1b-45c7-bd37-22b1cf542535
1140814 05 Jul 17:48:18.144 DBUtil INFO - getPuzzlesPublishedByTimeRange for getPuzzlesPublishedByTimeRange set: vox fromPublishTime: 1640995200000 toPublishTime: 1643673600000 # of crosswords: 26 DBtimeTaken(ms): 3 totalTimeTaken: 6 dbName: us_pmdb dbIp: 10.26.64.3
1140815 05 Jul 17:48:18.144 PuzzleHelper INFO - op: puzzlesFromDBOrDiskByPublishTime: 1640995200000 #_of_published_puzzles: 26
1140816 05 Jul 17:48:18.144 DBUtil INFO - getPuzzleCompletionStatusForUser for getPlayProgressForUser userId: a54aa6e0-0e1b-45c7-bd37-22b1cf542535 series: vox ids: ;PBvox_20220131_1100WNvox_20220129_1100AJRvox_20220128_1100WNvox_20220127_1100AOKvox_20220126_1100JGVox_20220125_1000PBv
1140817 05 Jul 17:48:18.150 DBUtil INFO - getFirstPublishTime for getFirstPublishTime sets: [vox] DBtimeTaken(ms): 4 dbName: us_pmdb dbIp: 10.26.64.3
1140818 05 Jul 17:48:18.151 LoggingFilter INFO - COMPLETED Request ip: 45.85.145.154, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/calendar-picker referer: https://cdn3.amuselabs.com/? with params set: vox embed: 1 maxCols: 1 month: 0 year: 2022 elapsedTime(ms): 27
1140819 05 Jul 17:48:18.275 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":true,"updateLoadTable":false,"series":"tny-weekly","id":"69ff9654","playId":"cc6e4f03-4d4b-4563-83ce-99357f5faf46","userId":"90cbbb61-733a-438c-b8dc-4a55b13337f9"} with params id: cc6e4f03-4d4b-4563-83ce-99357f5faf46 set: atlantic_20240701 rowForPlayMustExist true updated row
1140820 05 Jul 17:48:18.285 DBUtil INFO - status: success for addOrUpdatePlay playId: cc6e4f03-4d4b-4563-83ce-99357f5faf46 uid: 90cbbb61-733a-438c-b8dc-4a55b13337f9 series: tny-weekly id: 69ff9654 rowForPlayMustExist true updated row
1140821 05 Jul 17:48:18.285 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 9.634542 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140822 05 Jul 17:48:18.285 LoggingFilter INFO - COMPLETED Request ip: 216.164.52.175, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/tny/crossword?id=69ff9654&set=tny-weekly&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140823 05 Jul 17:48:18.395 DBUtil INFO - Get playSnapshot for getPlaySnapshot user_id: 1a296de5-b859-4f29-b706-fd0bae5e94fc series: vox id: AJRvox_20240705_1000 DBtimeTaken(ms): 1 dbName: us_pmdb dbIp: 10.26.64.3
1140824 05 Jul 17:48:18.396 JSPHelper INFO - PostScore received through sendBeacon postScoreJson: {"updatePlayTable":false,"updateLoadTable":false,"series":"tny-weekly","id":"69ff9654","playId":"962eff7f-650e-4a77-a9da-7f18fbcb28ff","userId":"1edc906e-6c3e-4de6-a6db-a13e0a97ad46"} with params id: 962eff7f-650e-4a77-a9da-7f18fbcb28ff set: atlantic_20240701 rowForPlayMustExist true updated row
1140825 05 Jul 17:48:18.396 JSPHelper INFO - PostScore Performance Log: getPlaySnapshotPostScore: 0.0 addOrUpdatePlayPostScore: 0.0 getMaxScorePostScore: 0.0 setLoadAsVerifiedPostScore: 0.0
1140826 05 Jul 17:48:18.397 LoggingFilter INFO - COMPLETED Request ip: 98.116.127.30, 34.149.208.254 url: http://cdn3.amuselabs.com/auth0/postScore referer: https://cdn3.amuselabs.com/tny/crossword?id=69ff9654&set=tny-weekly&embed=1&compact=1&maxCols=2&src=http%3A%2F%2Fwww.newyorker.com/
1140827 05 Jul 17:48:18.407 DBUtil INFO - status: success for addPlayToDBAndGetTimestamp playId: d000a502-eeb5-4a8f-985e-9f851161f0f uid: 1a296de5-b859-4f29-b706-fd0bae5e94fc id: AJRvox_20240705_1000 set: vox nRowsUpdated: 1 DBtimeTaken(ms): 12 dbName: us_pmdb dbIp: 10.26.64.3
1140828 05 Jul 17:48:18.414 DBUtil INFO - setLoadAsVerifiedAndUpdateId uid: 1a296de5-b859-4f29-b706-fd0bae5e94fc set: vox timestamp: 1720201686129 nAttempt: 1 UA: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/126.0.0.0 Safari/537.36 ip: 76.1
1140829 05 Jul 17:48:18.416 DBUtil INFO - Get playSettings for getPlaySettings userId: 1a296de5-b859-4f29-b706-fd0bae5e94fc series: vox puzzleType: CROSSWORD DBtimeTaken(ms): 1 dbName: us_pmdb dbIp: 10.26.64.3
```

Problem Statement

38 puzzles created by 16 constructors (0 new)

Puzzle type distribution

- CROSSWORD: 35 (92.1%)
- WORDSEARCH: 2 (5.3%)
- QUIZ: 1 (2.6%)

1. media24-netwerk24-quiz (account: null) (OLD_CREATOR) has created 90 so far

1) [VASVRA | Vr. 9 Augustus 2024](#) (QUIZ) Questions: **10** CreationStatus: **COMPLETE** creationTime:Aug 7 2024 12:59:48 UTC **Locale: af-ZA**

2. arbeitsversionphx (account: null) (OLD_CREATOR) has created 6 so far

1) [Standardrätsel H 10776](#) (CROSSWORD) Grid: **13x13** Words: **20** CreationStatus: ALL_CLUES_MISSING Fill: 63% **SpecialCells: 169** creationTime:Aug 7 2024 18:27:12 UTC
explicitPublishTime: Aug 28 2024 22:01:00 UTC Locale: de-AT Import: sourcePath:/unknown/enter_letters | **Oversize puzzle file (56KB)**

3. haldenphx (account: null) (OLD_CREATOR) has created 15 so far

1) [Standardrätsel H 55](#) (CROSSWORD) Grid: **13x13** Words: **20** CreationStatus: ALL_CLUES_MISSING Fill: 63% **SpecialCells: 169** creationTime:Aug 7 2024 16:27:45 UTC
explicitPublishTime: Oct 2 2024 22:01:00 UTC Locale: de-AT Import: sourcePath:/unknown/enter_letters | **Oversize puzzle file (56KB)**

4. media24-landbou-crossword (account: null) (OLD_CREATOR) has created 731 so far

1) [\(No title\)](#) (CROSSWORD) Grid: **9x9** Words: **20** CreationStatus: **COMPLETE** Fill: 66% creationTime:Aug 7 2024 06:34:24 UTC **explicitPublishTime: Aug 7 2024 22:00:00 UTC**
Locale: af-ZA Import: jpz/xml sourcePath:landbou-crossword_20240808.xml

5. media24-netwerk24-crossword (account: null) (OLD_CREATOR) has created 733 so far

1) [\(No title\)](#) (CROSSWORD) Grid: **6x8** Words: **18** CreationStatus: **COMPLETE** Fill: 91% creationTime:Aug 7 2024 06:30:24 UTC **explicitPublishTime: Aug 7 2024 22:00:00 UTC**
Locale: af-ZA Import: jpz/xml sourcePath:netwerk24-crossword_20240808.xml

Problem Statement

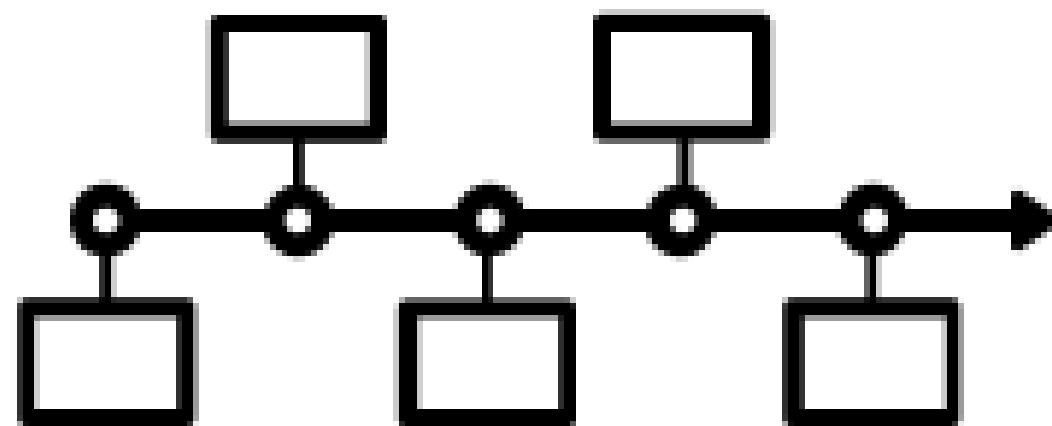
~15.9 GB of files representing **~322 million** logs every week

900+ series across **3** CDNs

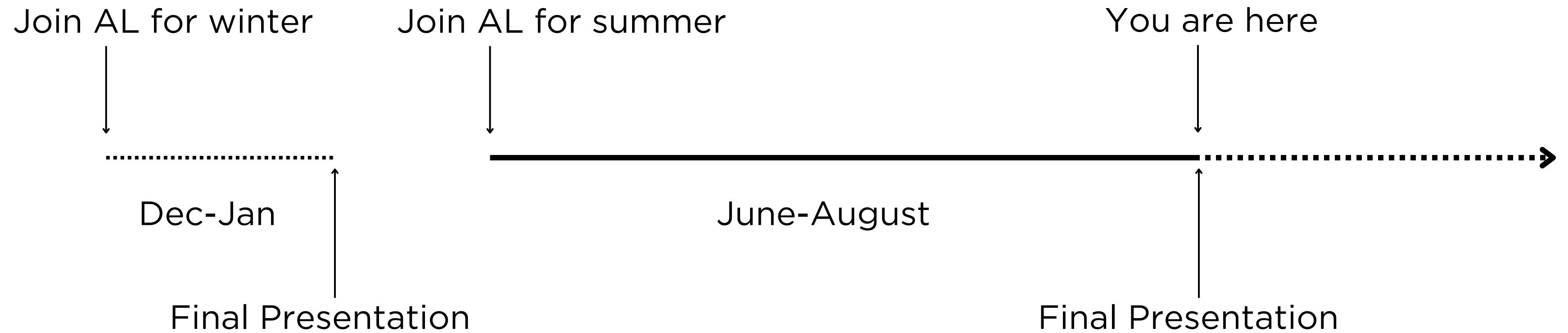
9 email reports generated by **6** report scripts

Process

Baby steps

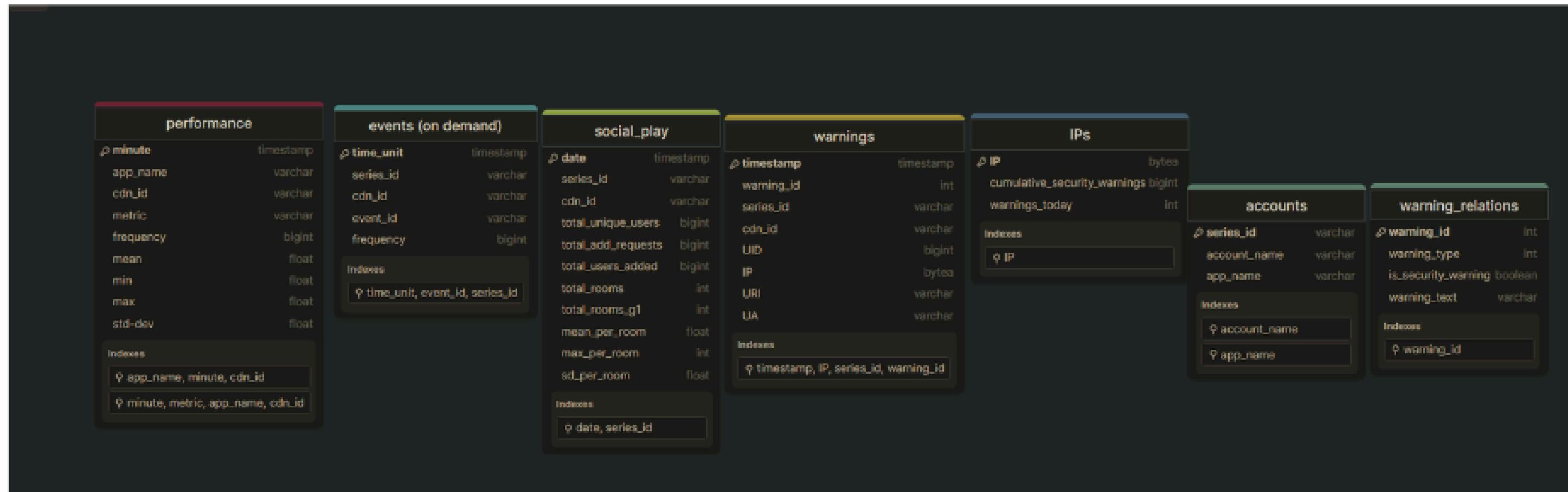


Timeline



TSDB (then)

Schema



TSDB (then)

Future Work

- 1 Add new fields and tables
- 2 Tailor visualizations to dev and business needs
- 3 Deploy

Timeline (then)

TSDB project starts



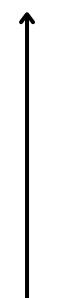
Start creating dashboards, only 2 - performance and warnings



Log report structure overhauled completely, all changes overwritten, nobody can find DB code



First iteration of the schema is created, logging scripts modified



DB and Dashboard are committed somewhere? but not deployed

Timeline (now)

TSDB project restarts

Recreate Superset, add puzzle play report and Biz/CS dashboards

Log report structure overhauled completely, all changes overwritten, deployment to be changed entirely

Recreate logging changes in the new structure, spin up DB with better schema

Deploy DB, set up SSH and VPN tunnels

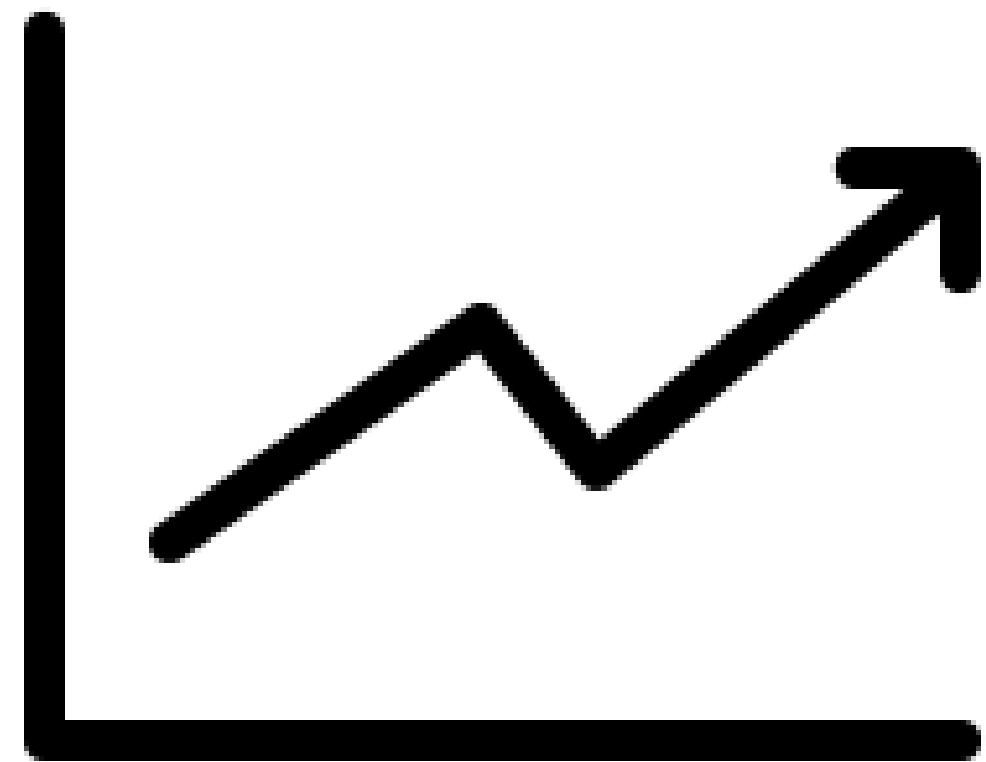
Commit log report changes to deploy, add dashboard features with feedback

Process

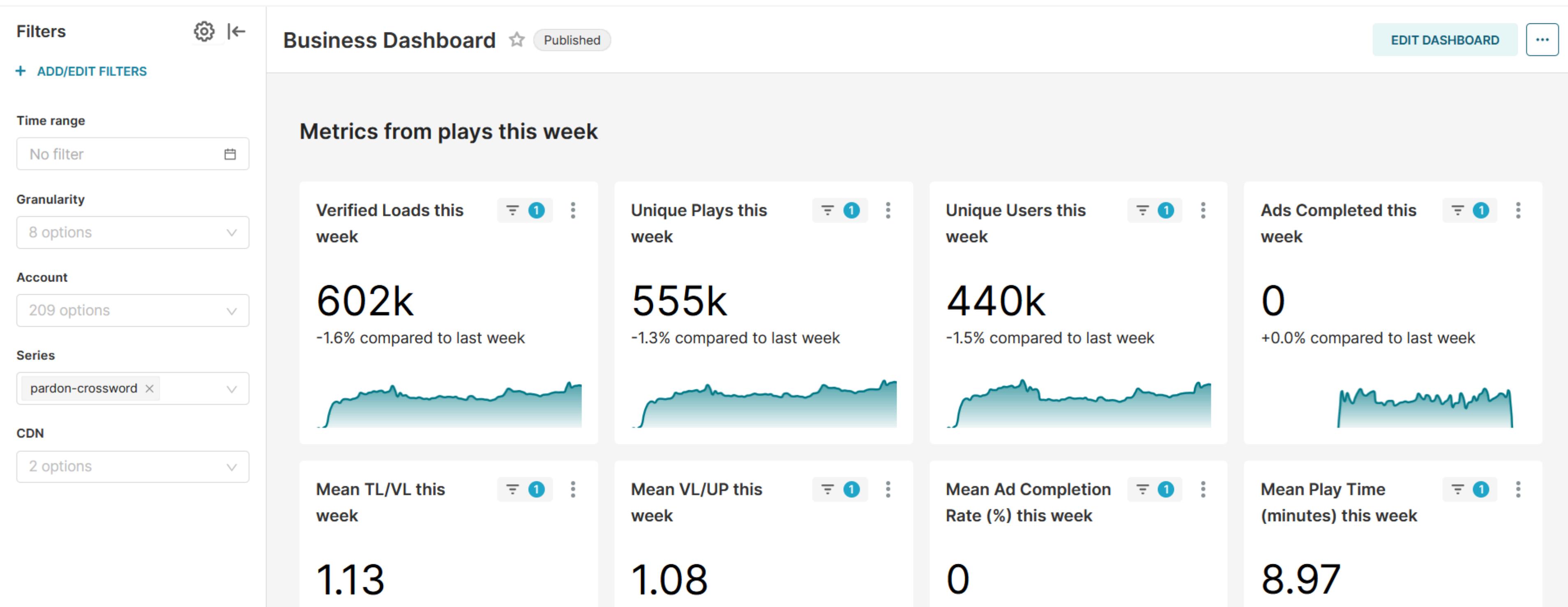
- 1 Design and implement a timeseries database in PostgreSQL, with aggregation and profiling extensions, in a Docker container
- 2 Parse and aggregate the data by building on top of the existing log reporting scripts and pushing to the database over VPN
- 3 Connect the database to Superset and create dashboards by generating queries for timeseries and rate-of-change information
- 4 Host the Pensieve and TSDB containers on Linodes, connect them over VPN, and expose Superset over SSH

Solution Spec: TADA

Timeseries Analytics DASHBOARD



The Solution



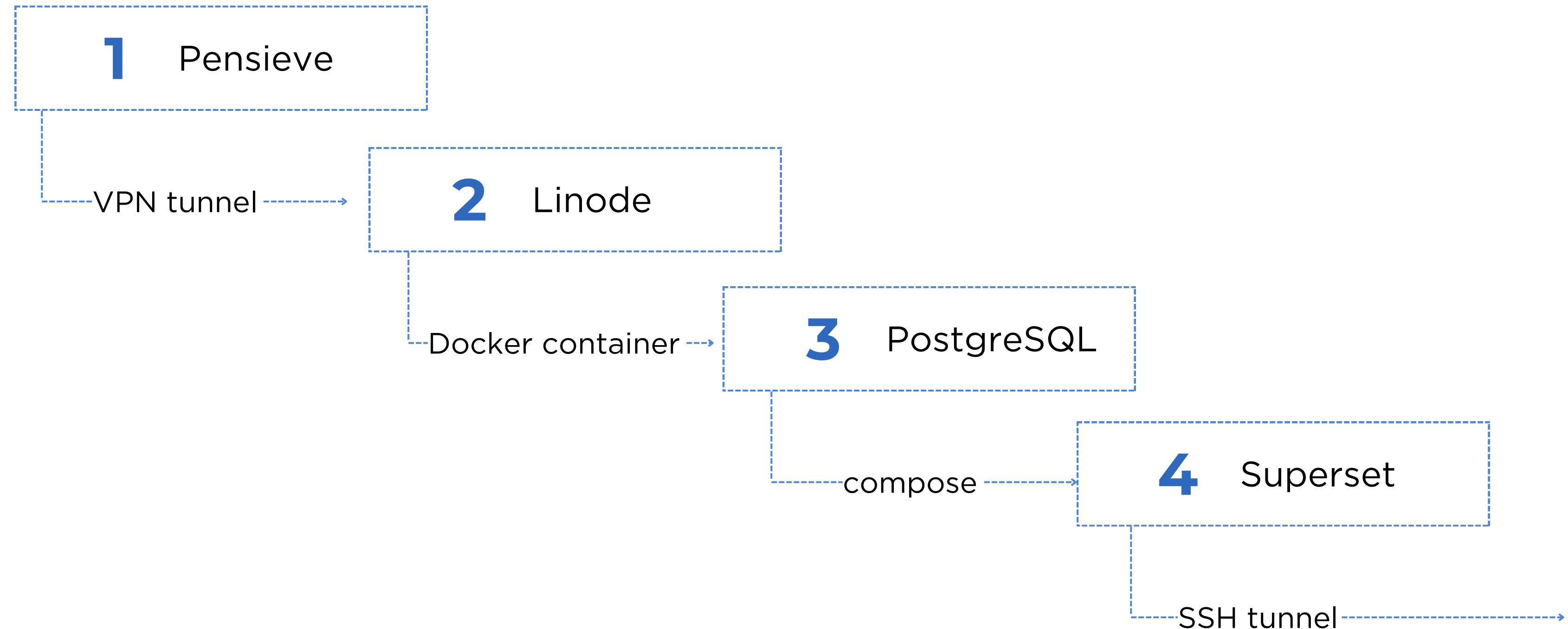
The Solution

The time series analytics pipeline automatically pushes every day's logs to a fast, indexed Postgres DB that can be accessed by Apache Superset to visualize weekly and historical data.

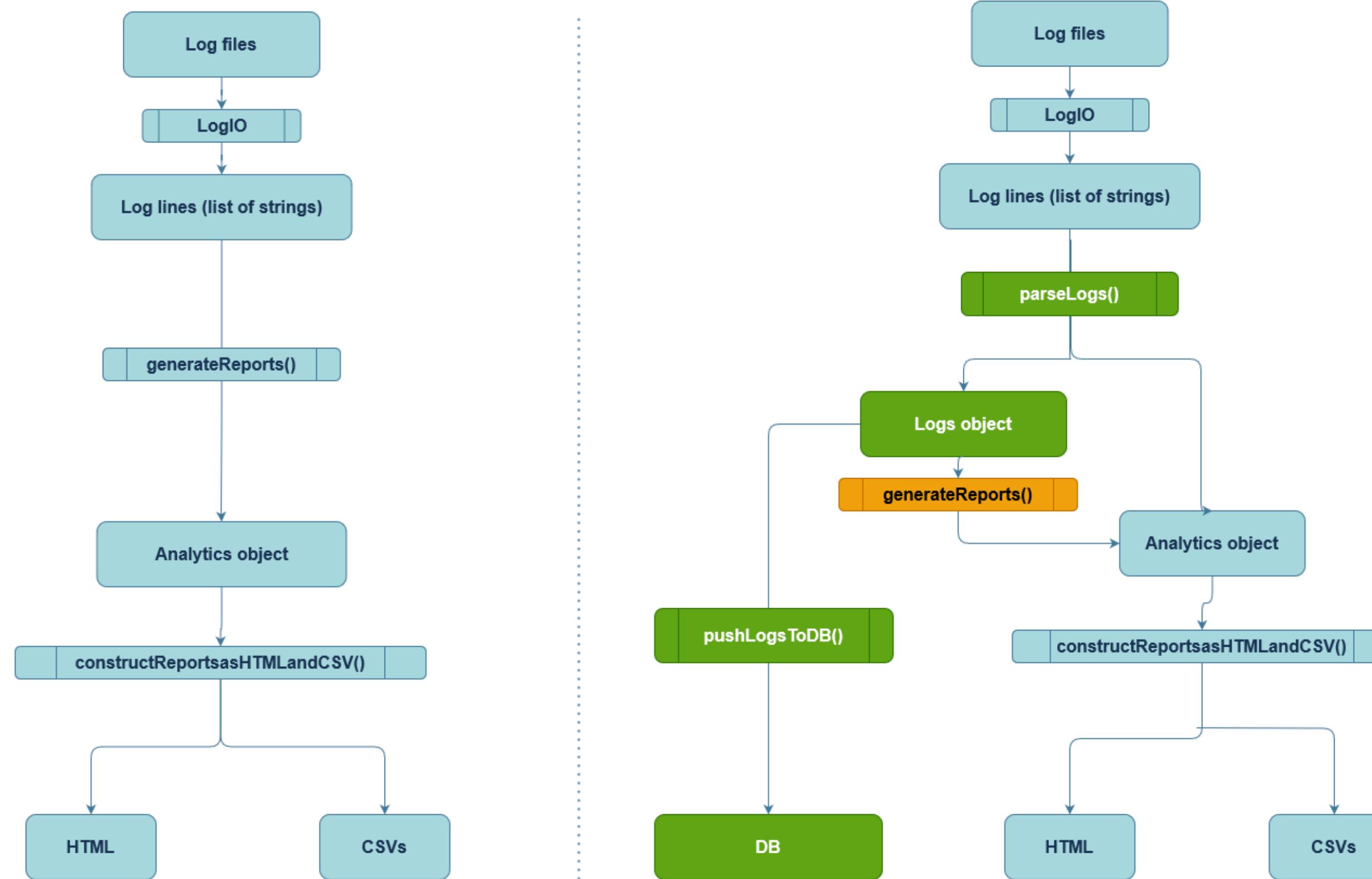
This solution:

- 1 Condenses key metrics from the emails into alerts and insights
- 2 Makes it much easier to track spikes and changes in the logs or plays
- 3 Allows historical investigation through aggregated logs
- 4 Reasonably performant (40k r/s) and scalable (20k r/MB)
- 5 Extensible as the logs and apps change
- 6 Accessible locally without exposure over the internet

Architecture



Logging (Pensieve)



Logging (Pensieve)

```
def pushToDB(logs: Logs, analytics: Analytics, appName: str):  
    """  
    Updates tables in the DB using the logs and analytics objects.  
    :param logs: An object of the Logs class. This contains the parsed logs used to construct the reports.  
    :param analytics: An object of the Analytics class. This contains the aggregated logs used to construct the reports.  
    :param appName: The name of the app for which the reports are to be generated. If not provided, appName is taken as empty string.  
    """  
  
    def pushPerformanceTable(performance_logs: dict, analytics_perf_logs: dict, appName: str):...  
  
    def pushWarningsTable(warning_logs: list, sec_warning_logs: list, bot_warning_logs: list):...  
  
    def pushRefererTable(referers: dict):...  
  
    def updateAccountsTable(warning_logs: list, sec_warning_logs: list, bot_warning_logs: list):...  
  
    if appName != "":  
        log.info(f"Updating logs to DB for {appName}.")  
    else:  
        log.info("Updating logs to DB.")  
  
    engine = sqlalchemy.create_engine(  
        f"postgresql://{{globalVariables.DB_PARAMS['user']}:{{globalVariables.DB_PARAMS['password']}}@{{globalVariables.DB_PARAMS['host']}:{{globalVariables.DB_PARAMS['port']}}/{{glob...  
  
    pushPerformanceTable(logs.performanceLogs, logs.analyticsPerformanceLogs, appName)  
    pushWarningsTable(logs.warningLogs, logs.securityWarningLogs, logs.botWarningLogs)  
    pushRefererTable(analytics.refererAnalytics)  
    updateAccountsTable(logs.warningLogs, logs.securityWarningLogs, logs.botWarningLogs)
```

Docker projects

python3.9

linuxserver/wireguard

amancevice/superset

postgres with pg_cron and pg_badger

linuxserver/wireguard

Postgres Schema

14th Aug '24

Visualization (Superset)

Filters

+ ADD/EDIT FILTERS

Time range No filter

Granularity Hour

Account 73 options

Series 325 options

CDN 2 options

APPLY FILTERS

CLEAR ALL

Warnings Dashboard Published

Metrics from warnings yesterday

Daily Warning Count 73.6k +38.4% since yesterday

Weekly Warning Count 73.6k

Daily Security Warning Count 1.29k -6.9% since yesterday

Weekly Security Warning Count 1.29k

Daily Change in Warnings by Series
Show 100 entries

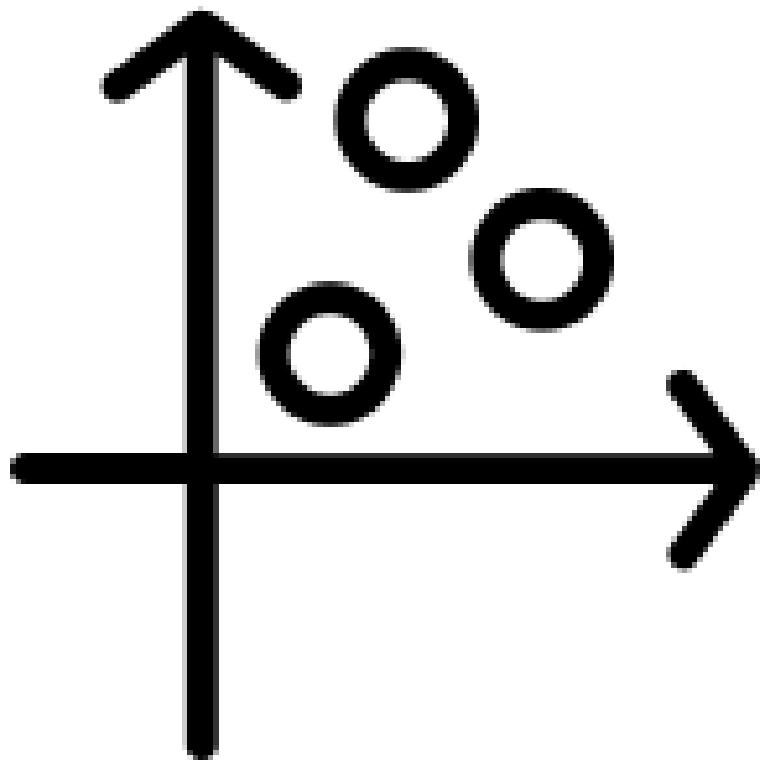
series_id	Change (%)	Change
pardon-crossword	152.49	7.67k
mw-missing-letter	210.62	2.88k
vox	67.37	2.16k
morningbrew-mini-crossword	2.96k	1.99k

Daily Change in Warnings
Show All entries

warning_text	Change (%)	Change
Exception while reading snapshot data from postScore.	301.35	889
Error playing audio.	19.89	368
Error in performing inputChanged.	106.77	189
AJAX error message - Error in posting scores, navigator.online is true.	34.29	143

Use Cases

For customers and internal



Common use cases

Weekly monitoring

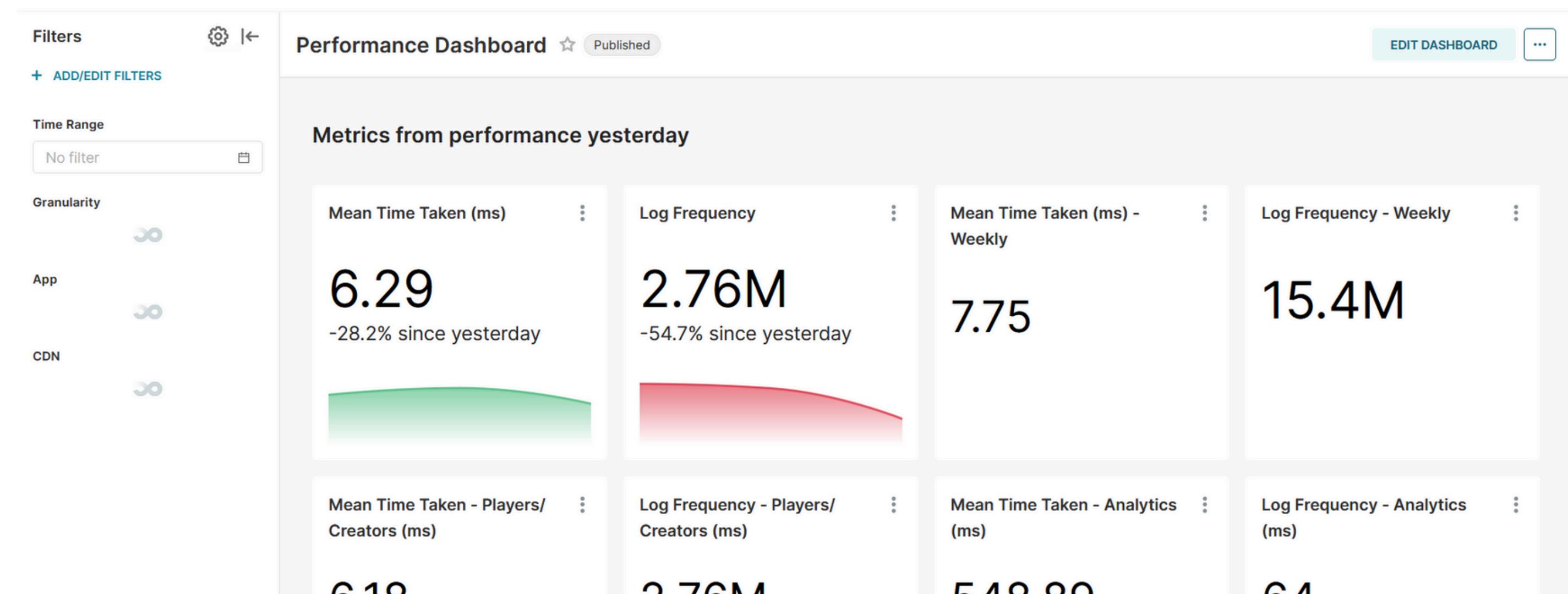
Incident investigation

Customer comparables

Longer-term context

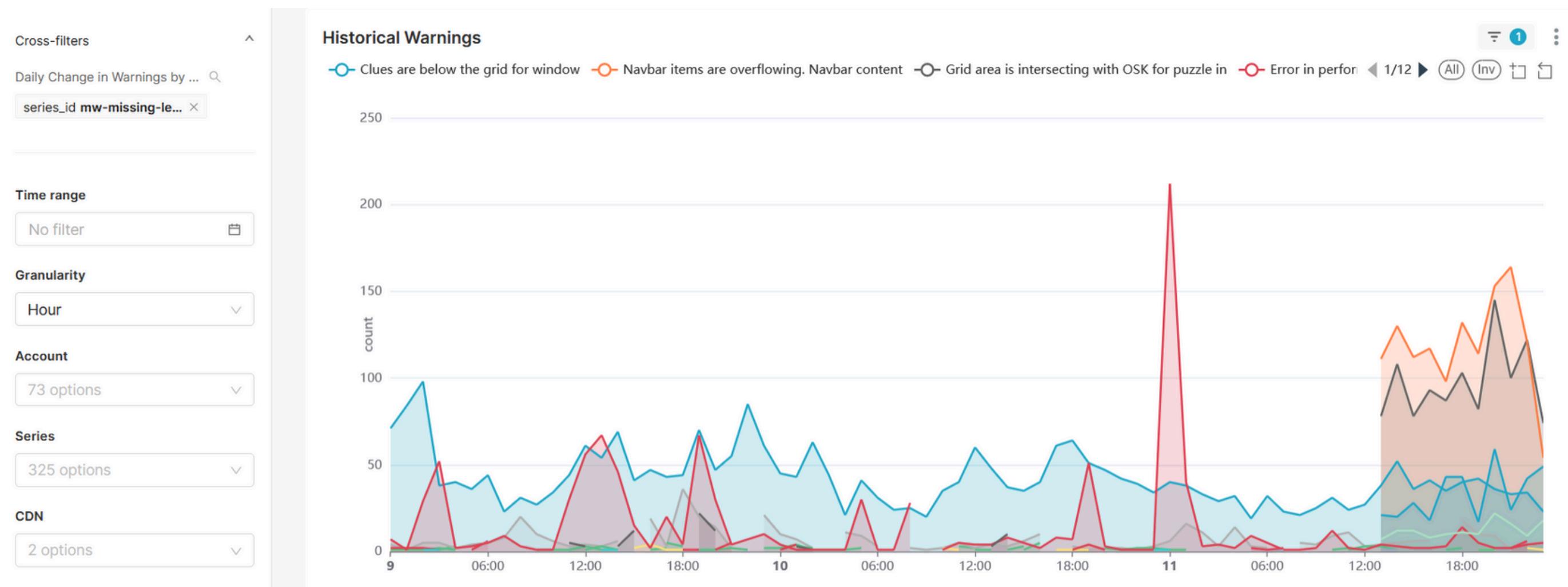
Weekly monitoring

To do routine checks, make sure performance is consistent, keep an eye on series of concern (e.g. with fewer than 500 loads), track enterprise customers



Incident investigation

To investigate outages, explore possible attacks and spikes in warnings, answer customer queries, or examine drops in plays



Customer comparables

To know what to expect for new customers and predict traffic for new and existing customers

Time range

Granularity

Account

Series

+ 6 ...

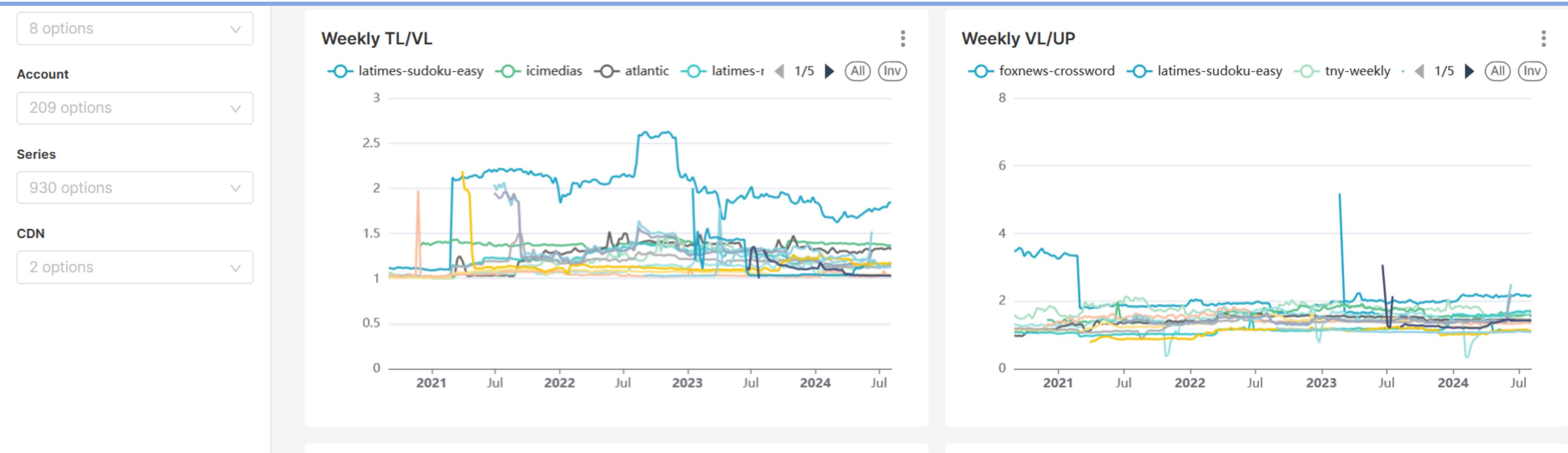
Business Numbers this week

Show All entries Search 9 records...

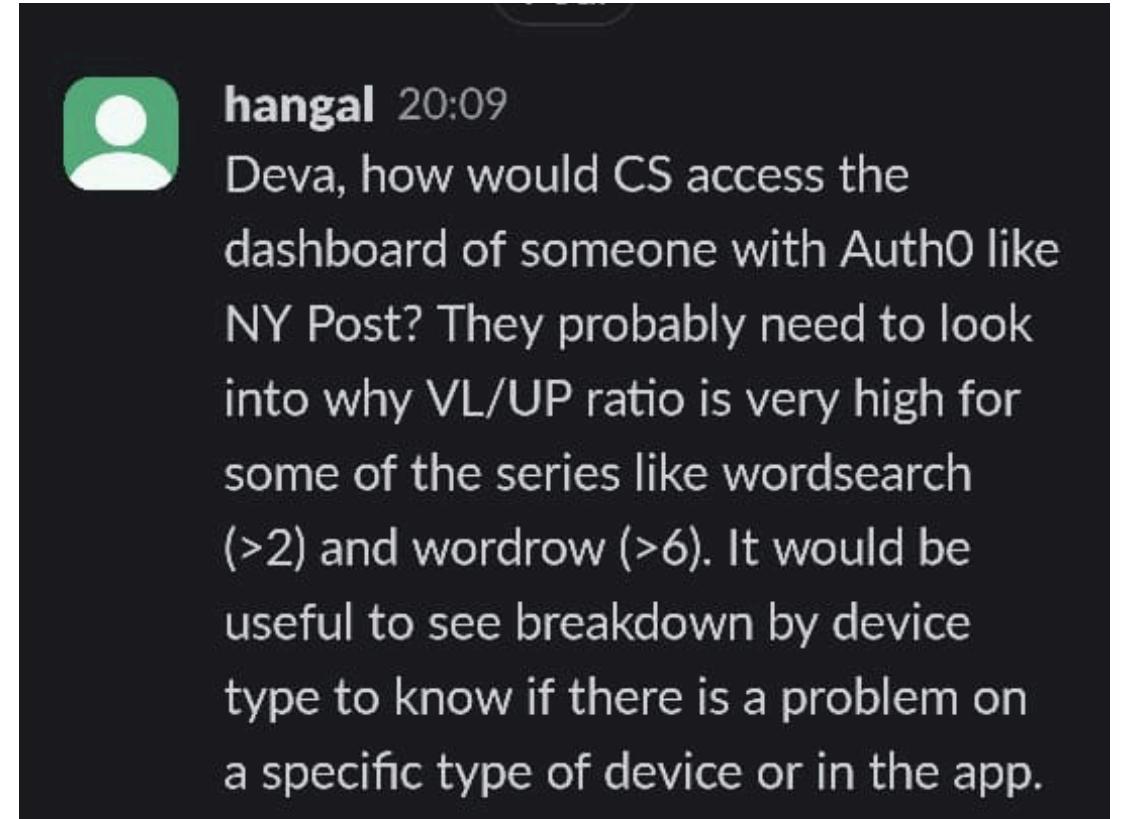
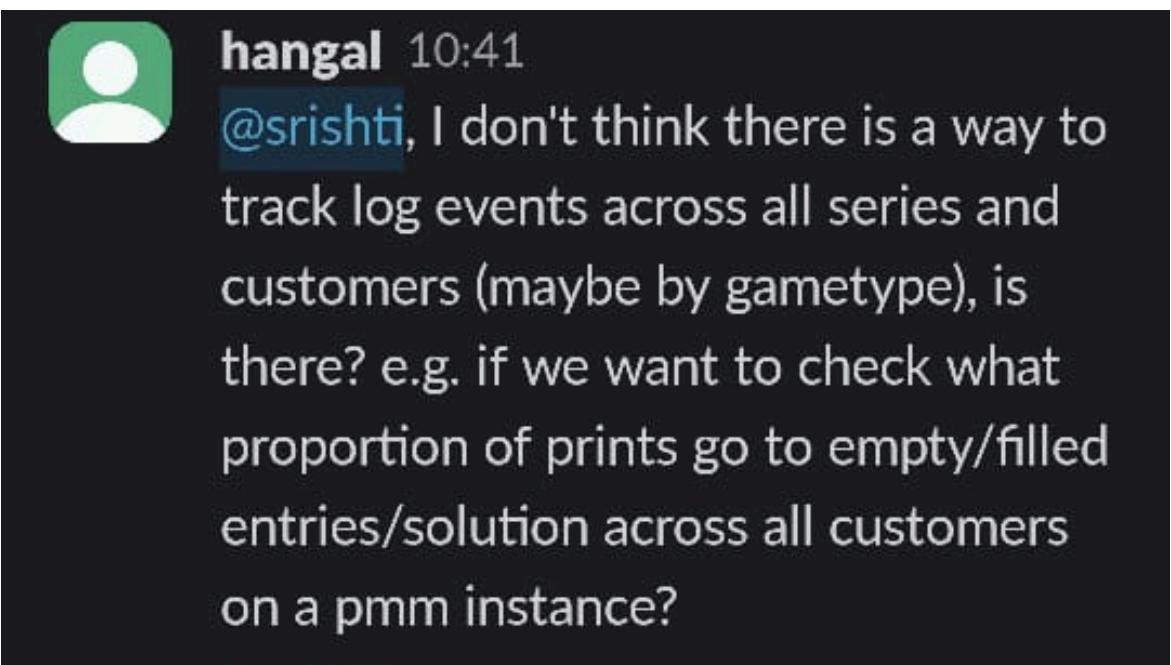
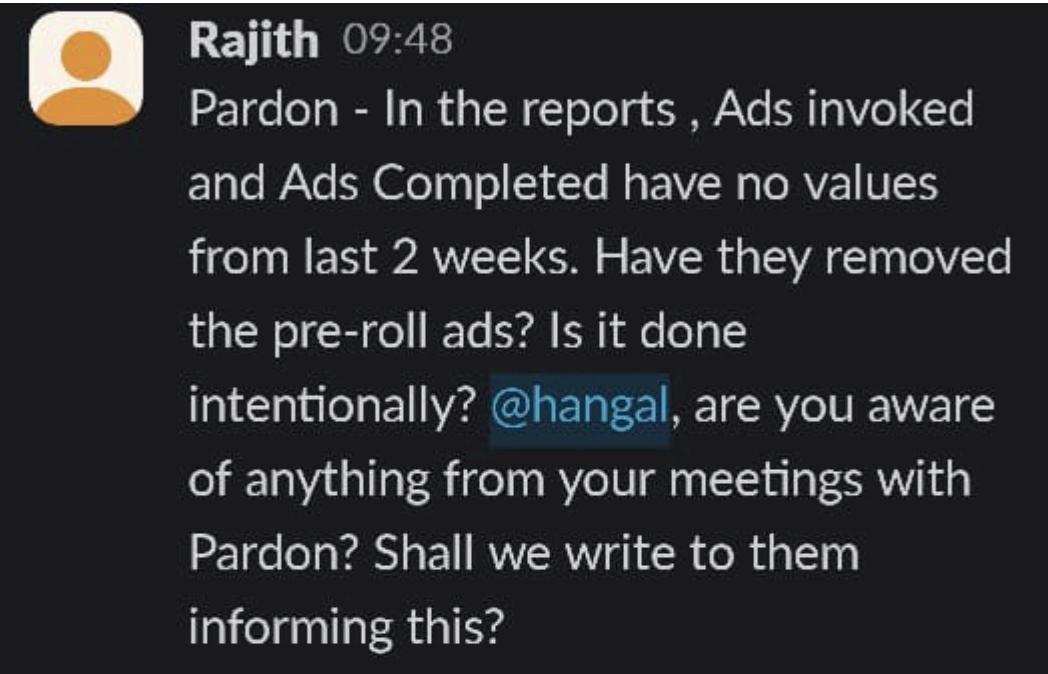
series_id	Total Loads	Verified Loads	Unique Plays	Unique Users	Ads Invoked	Ads Completed
boredpanda-wordrow	207	195	160	120	0	0
boredpanda-word-search	199	189	136	111	0	0
boredpanda-crossword	180	175	135	111	0	0
boredpanda-jigsaw	178	168	125	108	0	0
boredpanda-word-flower	151	142	93	85	0	0
boredpanda-kris-kross	119	104	93	88	0	0
boredpanda-picdoku	116	105	85	85	0	0
boredpanda-killer-sudoku	107	95	20	20	0	0
boredpanda-sudoku	106	97	74	68	0	0
Totals	1.36k	1.27k	921	796	0	0

Longer-term context

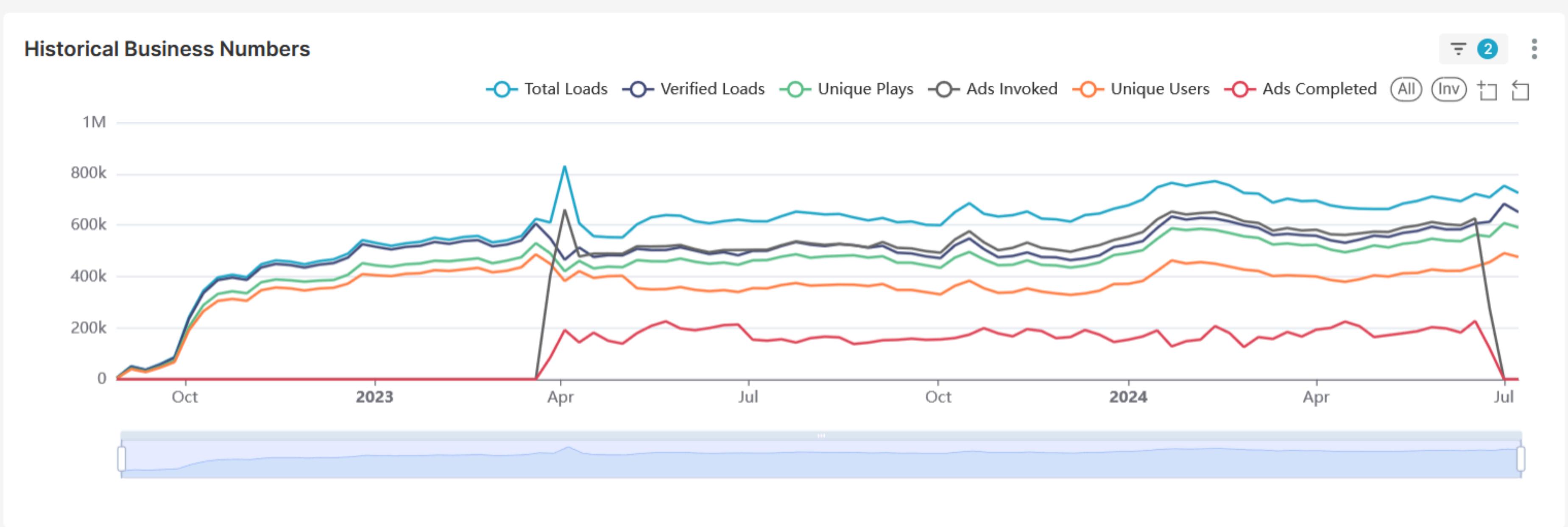
Track numbers and metrics over time to prevent creep, assess the impact of long term changes and events



Specific use cases

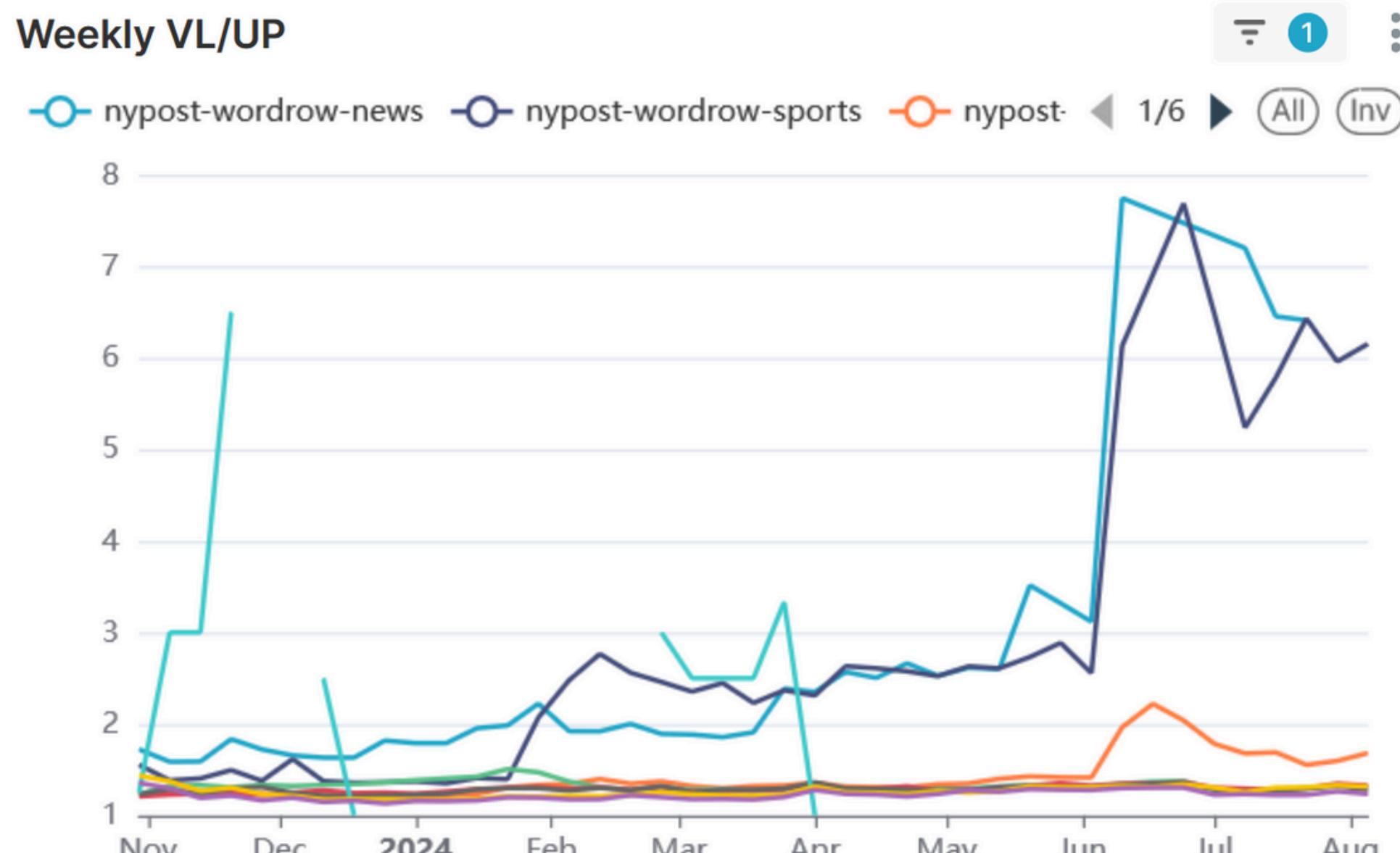


Specific use cases - Pardon

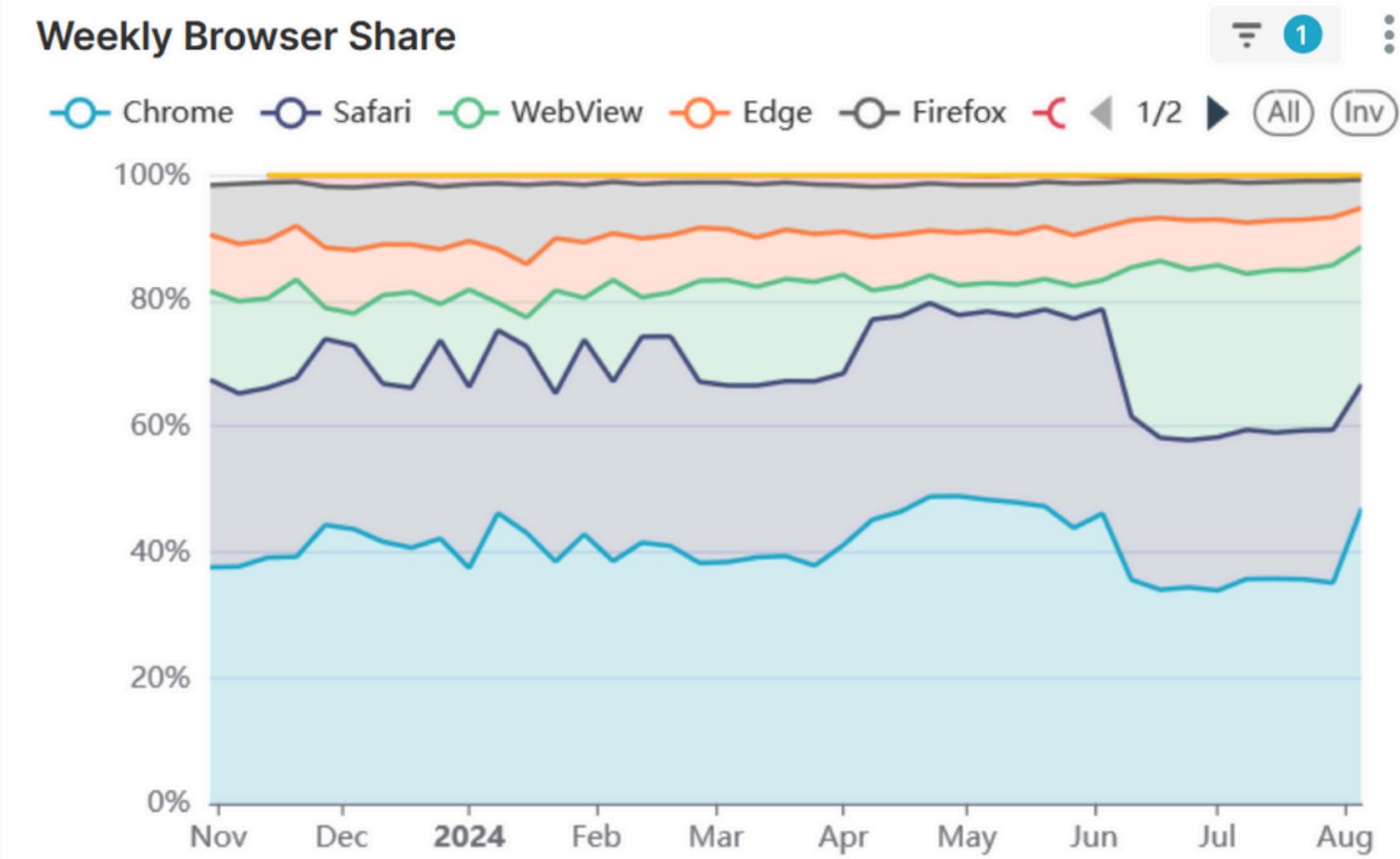


Specific use cases - VL/UP

Weekly VL/UP



Weekly Browser Share



Learnings

By leaps and bounds



What did I learn?

To write better SQL queries

```
WITH max_timestamp AS (
  SELECT MAX(timestamp) AS
  max_timestamp
  FROM puzzle_plays
),
current_week AS (
  SELECT
    series_id,
    cdn_id,
    SUM(total_loads) AS
  current_total_loads,
    SUM(verified_loads) AS
  current_verified_loads,
    SUM(ads_invoked) AS
  current_ads_invoked,
    SUM(ads_completed) AS
  current_ads_completed,
    SUM(unique_plays) AS
  current_unique_plays,
    SUM(unique_users) AS
  current_unique_users,
    AVG(mean_play_time) AS
  current_mean_play_time,
    SUM(plays_completed) AS
  current_plays_completed,
```

```
SUM(plays_completed_correctly) AS
  current_plays_completed_correctly,
  SUM(prints) AS current_prints
  FROM puzzle_plays, max_timestamp
  WHERE timestamp >
  max_timestamp.max_timestamp - INTERVAL
  '1 week'
  GROUP BY series_id, cdn_id
),
previous_week AS (
  SELECT
    series_id,
    cdn_id,
    SUM(total_loads) AS previous_total_loads,
    SUM(verified_loads) AS
  previous_verified_loads,
    SUM(ads_invoked) AS previous_ads_invoked,
    SUM(ads_completed) AS
  previous_ads_completed,
    SUM(unique_plays) AS
  previous_unique_plays,
    SUM(unique_users) AS
  previous_unique_users,
```

```
AVG(mean_play_time) AS
  previous_mean_play_time,
  SUM(plays_completed) AS
  previous_plays_completed,
  SUM(plays_completed_correctly) AS
  previous_plays_completed_correctly,
  SUM(prints) AS previous_prints
  FROM puzzle_plays, max_timestamp
  WHERE timestamp >=
  max_timestamp.max_timestamp - INTERVAL '2
  weeks' AND timestamp <
  max_timestamp.max_timestamp - INTERVAL '1
  week'
  GROUP BY series_id, cdn_id
)SELECT
  cw.series_id,
  cw.cdn_id,
  pw.previous_total_loads,
  cw.current_total_loads,
  pw.previous_verified_loads,
  cw.current_verified_loads,
  pw.previous_ads_invoked,
```

```
  cw.current_ads_invoked,
  pw.previous_ads_completed,
  cw.current_ads_completed,
  pw.previous_unique_plays,
  cw.current_unique_plays,
  pw.previous_unique_users,
  cw.current_unique_users,
  pw.previous_mean_play_time,
  cw.current_mean_play_time,
  pw.previous_plays_completed,
  cw.current_plays_completed,
  pw.previous_plays_completed_correctly,
  cw.current_plays_completed_correctly,
  pw.previous_prints,
  cw.current_prints
  FROM current_week cw
  LEFT JOIN previous_week pw
  ON cw.series_id = pw.series_id AND cw.cdn_id =
  pw.cdn_id;
```

What did I learn?

To write config, fix version issues, and hack workarounds

From cobbling together all the moving parts:

Bash

Superset

SQL

pg_cron

Postgres

Docker

Wireguard

Pandas

Python

What did I learn?

To drive projects forward and take initiative

- Yesterday
-  **karnav** 09:51
Hi prof, I've committed the log coding
- 2 Aug
-  **karnav** 09:47
Hi prof, what would be a good time to
- 4 Jul
-  **karnav** 11:01
Hi prof, would meeting today be
- 9 Aug
-  **karnav** 12:06
Hi prof, the tracking that SH shared on
- 30 Jul
-  **karnav** 10:41
Hi prof, just confirming that this'd be
- 3 Jul
-  **karnav** 10:57
Hi prof, would you be available to
- 6 Aug
-  **karnav** 14:19
Hi prof, would you have some time
-  1 
-  **karnav** 12:11
Hi prof, I'm committing the database
- 12 Jul
-  **karnav** 12:32
Hi prof, just confirming, did we say
- 26 Jun
-  **karnav** 14:24
Hi prof, would meeting today be
-  **karnav** 17:01
Hi prof, this is the list of all the
-  **karnav** 11:24
Hi prof, would you be available for a
-  **karnav** Hi prof,
-  **karnav** Hi prof,
-  **karnav** Hi prof,
-  **karnav** Hi prof,

What did I learn?

The name of my roommate next year



brothers



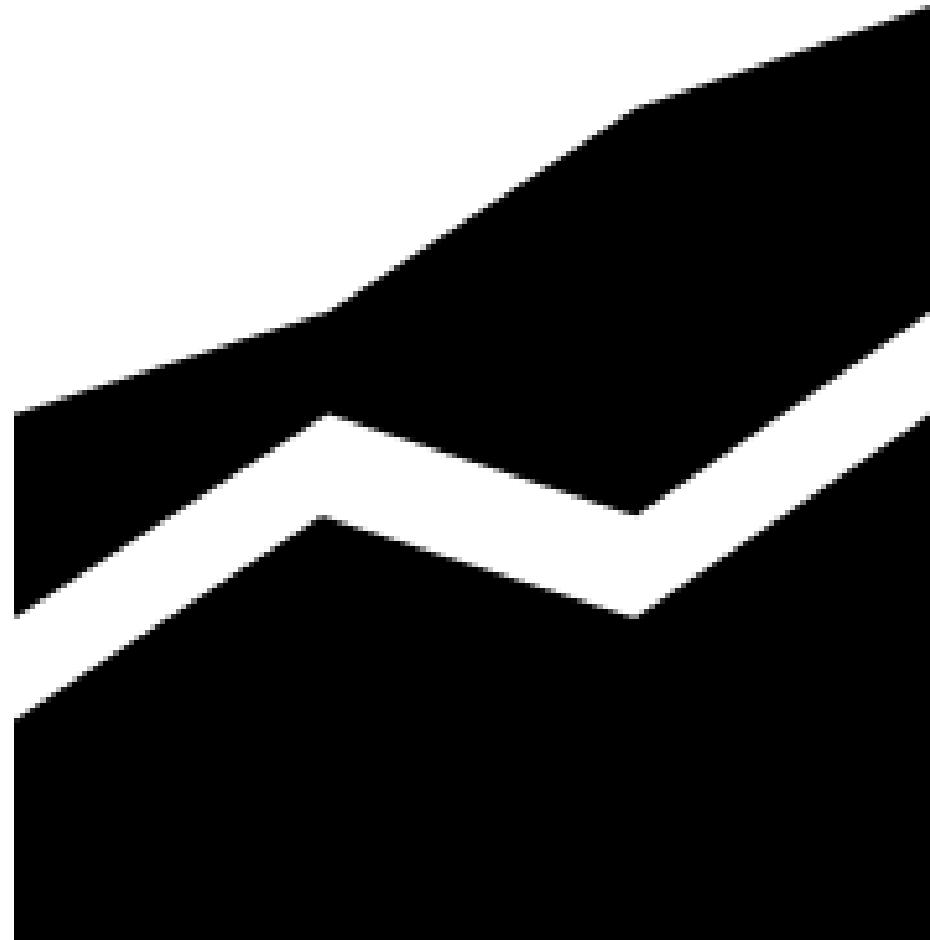
AL



roommates

Challenges and Future

A long and treacherous path forward



Challenges

Getting all the systems to work together

Following good practices in a complicated design

Upgrading infra to modern versions

Finding meaningful but efficient ways to graph data

Future Work

Rebuilding the log ingestion pipeline

Will the full pipeline ever get connected and running?

Superset inside PMM?

Finetuning dashboards to internal needs

Future Work

Integrating Rajith's puzzletype and paywall mapping

UID/user journey metrics?

Automating profiling and more efficient indexes

Thanking everyone at AL for a wonderful summer

Thank you!