Time Is on My Side: A Case Study of Query Sharing using Alation

Naveen Kalyanasamy, Andrea Levy, Aaron Kalb

Alation Inc.
Redwood City, CA
Info @ https://www.alation.com/

Abstract: We study how the Alation platform assists data-driven companies in saving time and effort through collaboration among their employees. We quantify the extent to which an organization has achieved some of these benefits through a specific case study of query sharing. For analysts writing SQL queries, we examine time saved by finding and reusing queries written by peers in Alation. The time savings and benefits of using the platform extend well beyond this area but are not covered in the present work.

Three quarters of organizations using Alation share queries across individuals, and this case study highlights the impressive time savings that one organization achieved. Using a sample of anonymized usage data, we quantify the effort users have saved from query reuse. In this organization, which has 572 users finding and sharing queries in Alation, we find more than 325 work days of time savings in one year.

Keywords: Analysts, Querying Databases, Collaboration, Data Culture

1. Background

For organizations striving to be data-driven and aiming to promote a data-culture, enabling and facilitating collaboration within and across analyst and data management teams is key [1]. This collaboration can contribute to enhanced and accelerated net output of individual teams. Collaboration among peers within the organization saves time and effort by enabling them to learn from and help each other. This information exchange also helps break down knowledge silos. Significant productivity improvements have been seen (as measured by query output rate) in analysts who are provided data platforms to work in that are collaborative in nature [2].

This case study investigates one organization's time savings and efficiency gains in the analysis process through sharing and reusing queries in Alation – a software platform which (among other things) enables collaboration across analytics teams. In organizations using the Alation platform, we find that sharing queries is common. Almost 65% of organizations[1] using Alation to query data have users sharing queries.

2. Theory

Extensive research has been performed on the analysis process and how analysts find, understand, trust, and process data to answer business questions and generate reports. These studies [3][4][5] conclude that an integral part of this process is the analysts' data preparation, which includes the following: discovery of the right data, understanding of the data, and wrangling of the data into the desired format. The analysts then model the data and summarize it in their final query.

Alation's data catalog enables analysts on the platform to search and find queries that other users have composed. Users are able to find and repurpose their peers' queries for their own goals and avoid building the query logic from scratch. Thus they end up saving time by leveraging the effort their peers contributed in following the process mentioned above to support a previous analysis.

Alation also provides the "Query Forms" feature, which facilitates query authors sharing their queries. These query forms enable their peers, who aren't familiar with writing code, to run the query and see the results without having to edit or even enter the coding environment. Query forms can also be configured to accept user-input values (e.g. to alter the filters of a query), which further increase the power and relevance of these authors' queries. We find that 73% of organizations sharing queries in Alation use query forms.

Alation assists users in borrowing from other users' analytical experience through other means as well, such as query cloning and crowd-sourced contextual information from peers about data objects. There is other potential for re-use apart from querying, such as discovery of relevant business intelligence reports, instead of having to recreate them from scratch. Even though there are broader benefits to using the Alation platform, the aim of this case study is to quantify and demonstrate the real-world impact of analysts sharing and re-using queries and to measure the amount of time and effort that they have saved by studying users' actions on the Alation platform.

3. Data

We use anonymized data logs summarizing when queries are executed and edited to determine when queries are developed and accessed. Individual users and data objects cannot be identified through these logs. Additionally, text of the queries are not available for review and cannot be used in the study. In this case study, we explore the query-sharing behavior of 572 users in a large e-commerce company where employees write a lot of queries. We look at 12 months of query edit and execution data to

[1] Organizations reporting usage data to Alation

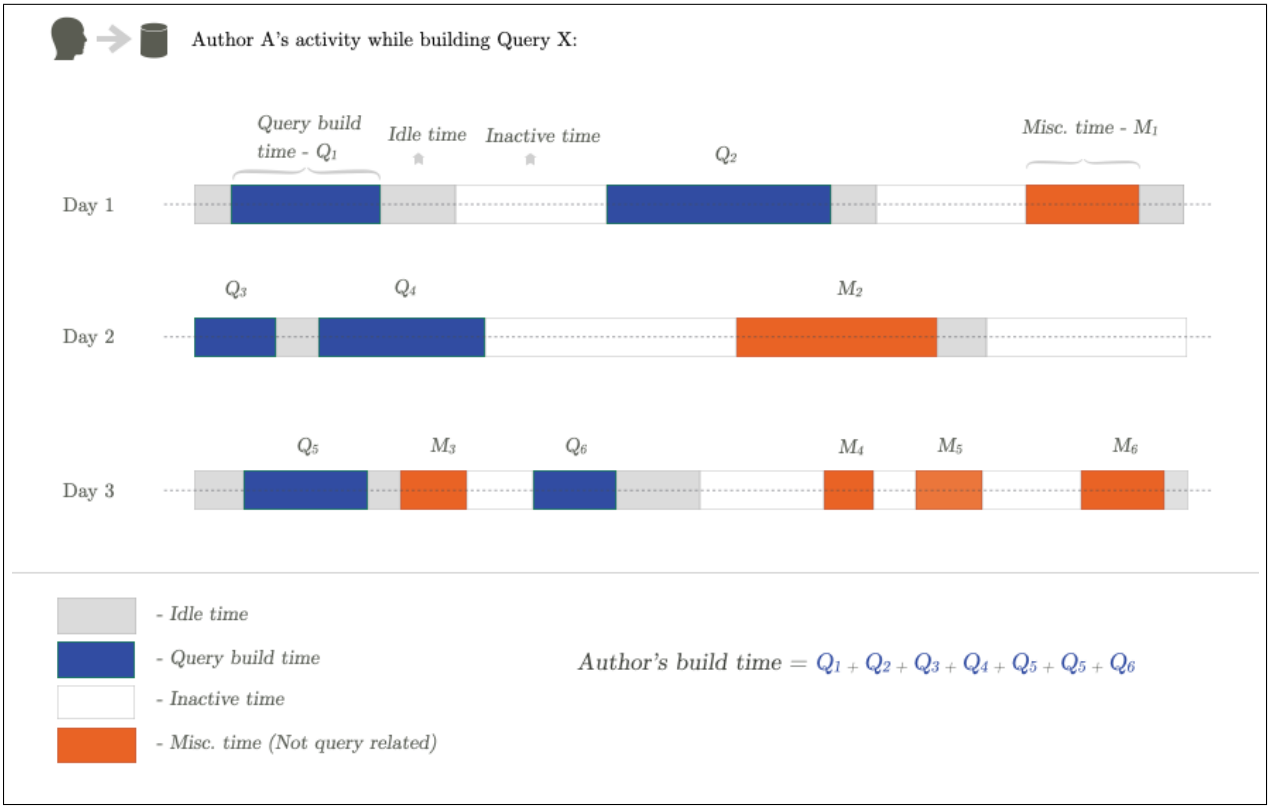

Fig. 1: Query build time calculation example

determine the annual impact of re-using queries.

4. Method

4.1. Non-author Executions

We started with a base list of users who found and ran queries that they did not write, along with the query identifier(s) and the timestamp of when the queries were executed. We can then reasonably theorize that the time saved by the user who re-used a query would be at least as much as the time spent by the author on building that query. It is worth noting that the time savings from reuse, is likely greater than just the presumed savings of the original author's compose time. In many cases, the original author possesses better domain knowledge or technical skills and in these scenarios the time for the expert to devise the query has been proved to be significantly lower than a beginner [2]. With the data and information we have in hand however, we treat the author's build time as a conservative estimate of the lower bound of time savings a reuser can achieve.

4.2. Author Build Time

For each query that was executed by someone who did not write it (a non-author), we calculate the time it took the author to build out that query. The build time of a query is the amount of time the author spent editing and working on the query in Compose (the query editing tool in Alation).

We investigate each session of time when the au-

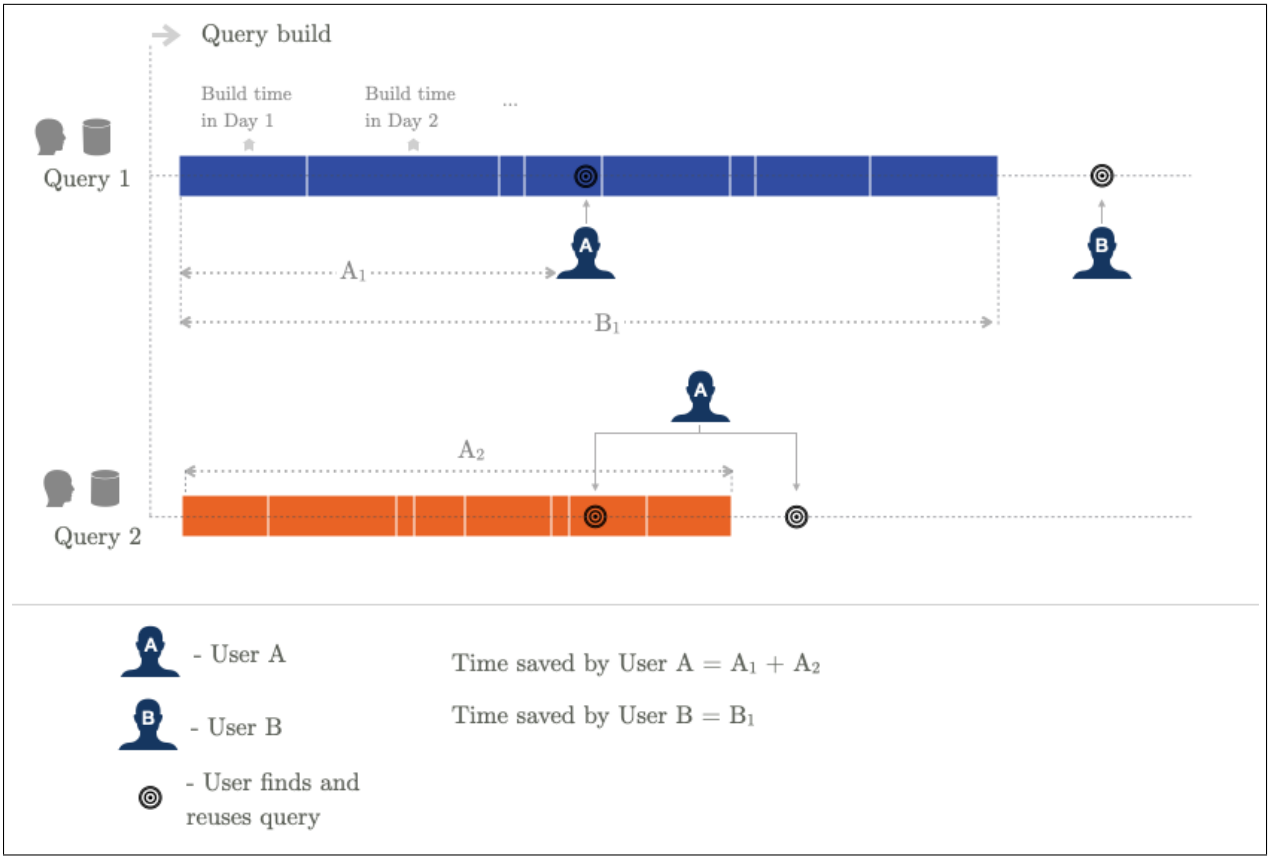

Fig. 2: Attribution of time-savings to users running others' queries

thor is active on the catalog. Supplementing these sessions with data generated when the author visits, edits, saves, and executes a query, we isolate the windows of time the author spent on building the query.

The total build time of an author for query q can be calculated as follows:

$$B_q = \sum_{i=1}^{M} Q_i \tag{1}$$

where Q_i is each window of time the author spent writing the query.

Fig.1 shows an example of an author's query building timeline and how the build time is calculated. Query build time is calculated by adding up the time the author spent writing the query (Q_1,...) and ignoring the time spent on miscellaneous activities (M_1,...) unrelated to the query.

4.3. Time Saved By Re-user

For each query a user runs that they did not build, the time that the author spent building the query (Section 4.2, Eqn.1) is assigned to that user as the time they saved by avoiding building the query from scratch. The time saved by each user is aggregated by adding the time saved for each query they reused. Note that if a user runs the same query multiple times, the time saved by running that query is only counted once. Were that user to write the

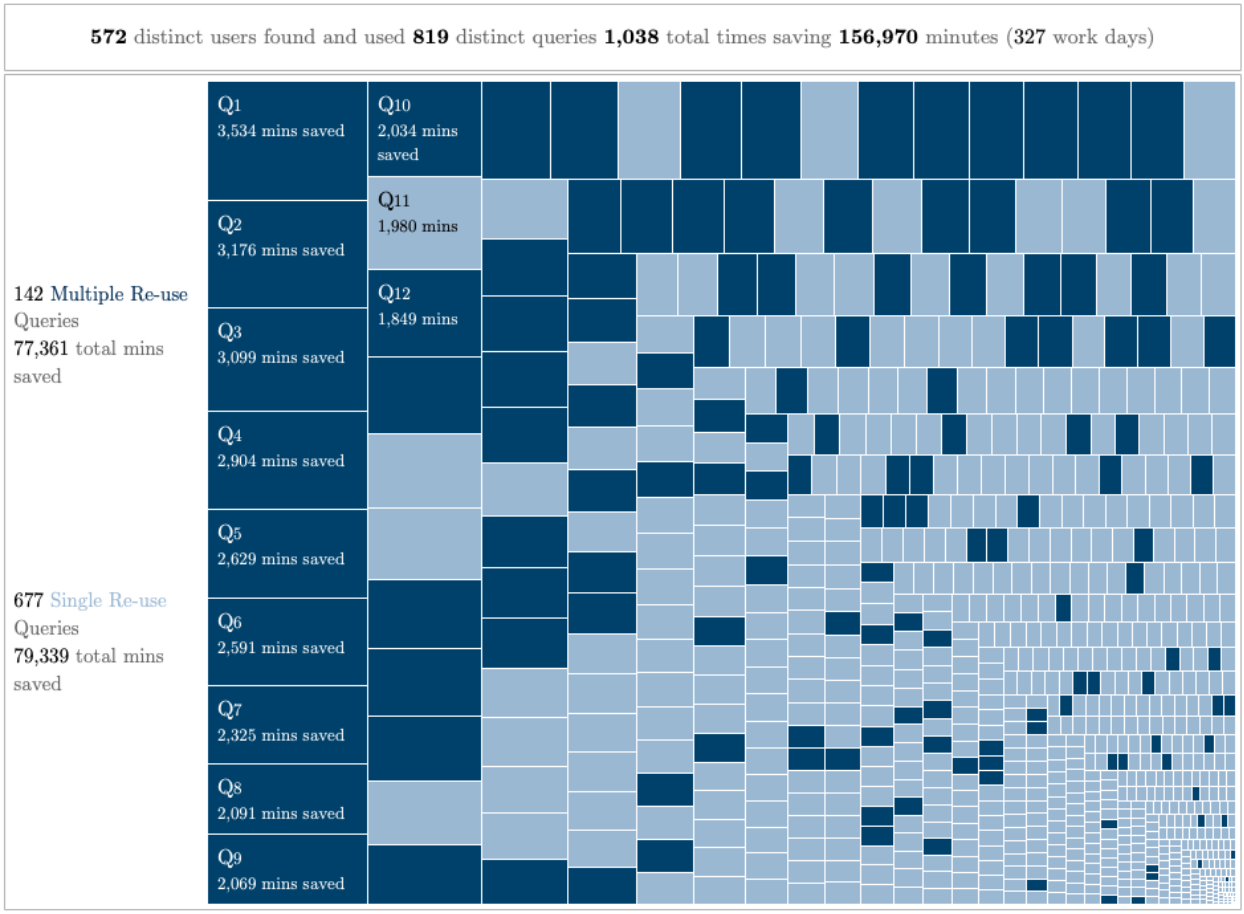

Fig. 3: Break down of 327 work days of savings by query savings per query. Queries that are re-used by multiple users, as well as queries reused by only one other person both contribute to significant time savings.

query themselves, they presumably would write it once and use it multiple times.

Total time saved by a user in a time period is defined as the sum of the build time B_q of all queries re-used q in the time period.

$$S = \sum_{q=1}^{N} B_q \tag{2}$$

The build time B_q for each query is calculated by Eqn. 1. Figure 2 represents this calculation pictorially. In Fig 2, total time saved by User A is $A_1 + A_2$. A_1 is the time Author 1 spent building Query 1 which user A re-used. It includes only the build time that took place prior to User A executing Query 1. A_2 is the time Author 2 spent building Query 2, which user A re-used twice - during and after the query build. User A gets credit only for the second execution (no double counting) since they would derive the most benefit from the query after the build is complete. Build times A_1 and A_2 would be calculated as a result of Eqn. 1.

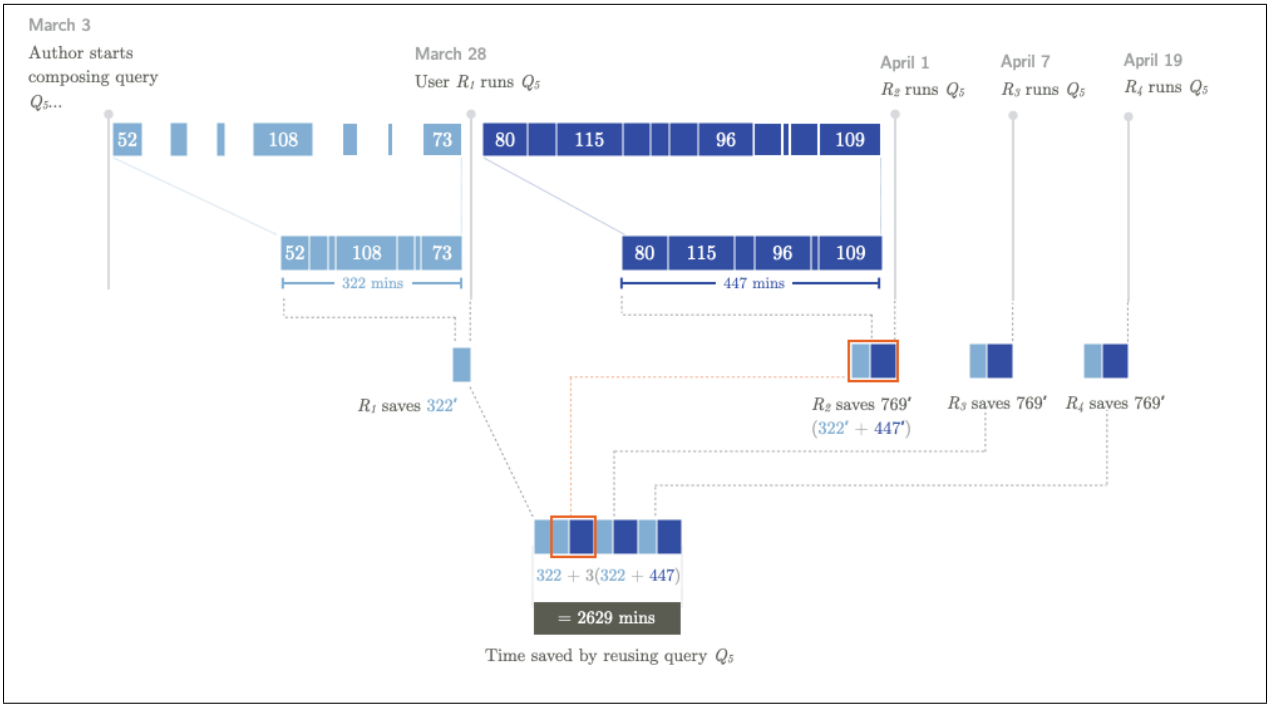

Fig. 4: Breaking down a specific query's journey to time savings

5. Results

Using the above methodology, we can aggregate the time each user has saved through non-author query executions in Alation. We can summarize across all queries in an organization to calculate the organizational time savings from query re-use, as shown in Fig. 3.

This organization saved 327 work days over the course of a year from users finding and reusing their peers' queries. These savings came from a mix of queries that were re-used multiple times ("Multiple re-use queries" in Fig. 3), as well as queries that were found by only one other peer ("Single re-use queries" in Fig. 3). Multiple re-use queries contributed 162 work days (77,631 minutes) of savings; single re-use queries saved 165 work days (79,339 minutes) in total.

Table 1 shows the 12 queries that have saved the most time at this organization (also annotated in Fig. 3). This list contains a combination of queries that took the author a lot of time and effort to write and were used by a few other people and also queries that took relatively less time to write but many of peers found useful. Query 1 in the table, for example, took the author 2,001 minutes to write across 211 sessions (Q_i in Section 4.2) over a span of 48 days and was found and used by 2 other peers. Query 9 on the other hand took only 236 minutes to write over a span of 10 days but 9 other peers saved time when they found and reused it for their purposes. We also notice that Query 11, which took the author 1,985 minutes to build over 70 sessions across 22 days, still saved substantial time for the one other user who utilized it.

The time saved by individual users who find and re-use a query may differ depending on what point

Query	Total time saved (mins)	Users who found and reused	Authoring time (mins)	Number of sessions	Days to write (span)
1	3,534	2	2,001	211	48
2	3,176	2	1,588	191	64
3	3,099	2	2,428	237	60
4	2,904	3	1,282	151	57
5	2,629	4	769	96	34
6	2,591	10	316	67	31
7	2,325	3	775	64	8
8	2,091	7	333	65	45
9	2,069	9	236	44	10
10	2,034	2	1,568	219	67
11	1,980	1	1,985	70	22
12	1,849	2	1,159	215	104

Table 1: Top 12 time-saving queries and the author's effort in composing them

in the author's query development they ran the query (see Fig 2). If we take a closer look at one of the queries that was found and leveraged by multiple users across this organization, we can break down the journey of the query from its inception to the time savings for the overall organization. Fig. 4 is a visual depiction of the development and usage journey of Query 5 in Table 1. It summarizes the time it took the author of Q_5 to construct it across multiple "build" sessions, which add up to 769 minutes (calculated using Eqn. 1). We can also observe the time saved by each user (R_1 to R_4), who found and used it, depending on what point in the query's development they ran the query. Adding up these times gives us nearly 44 hours saved from just one query re-used in this organization.

6. Summary

The Alation platform contributes to increased analyst productivity through enhanced data discovery, reduced query composing time and collaborative analytics. In this case study, we see that:

- Alation's collaborative platform assists analysts in saving effort by enabling them to discover and re-purpose their peers' queries and avoid building them from scratch.
- One organization leveraging Alation to re-use queries has saved nearly 327 work days across 572 users on the platform over the course of a year.

7. Generalizations

This analysis depicts the time savings and value that this specific organization has derived from one key feature area of Alation. We recognize that all organizations may not work in the same scale or scope of this one particular case and the results may vary depending on how many users/analysts are in the organization and the time they spend on analytical efforts. Among the organizations in our dataset, almost all see benefits from re-using queries.

- Over 65% of organizations have at least 10 users collaborating, finding, and reusing

queries; these organizations save an average of 37 work days per year

- 37% of the organizations have 20 or more users collaborating and save an average of 60 work days per year.

This distribution of realized savings across organizations using Alation showcases that while not all organizations see the magnitude of time savings demonstrated in this case study, sharing queries can result in significant time savings, even in less extreme cases.

The results in this case study are also not exhaustive of all the returns an organization can derive from Alation but do quantify a lower bound of time saved by actual organizations using one component of the Alation Data Catalog. Additional information on the capabilities of the platform can be obtained from the website [6].

References

[1] Gartner Inc. *Building a data-driven organization.* URL: https://www.gartner.com/smarterwithgartner/build-a-data-driven-organization/. (accessed: 11.15.2019).

[2] Y. Yin et al. *Learning by Doing versus Learning by Viewing: An Empirical Study of Data Analyst Productivity on a Collaborative Platform at eBay.* URL: https://www.kellogg.northwestern.edu/faculty/vanmieghem/htm/pubs/Learning-by-Viewing-at-eBay-cscw18_article193.pdf. (accessed: 11.15.2019).

[3] Sean Kandel et al. *Enterprise Data Analysis and Visualization: An Interview Study.* URL: http://vis.stanford.edu/files/2012-EnterpriseAnalysisInterviews-VAST.pdf. (accessed: 11.15.2019).

[4] Daniel M. Russell et al. *The Cost Structure of Sensemaking: Analysis of Three Case Studies.* URL: https://pdfs.semanticscholar.org/413f/037a150eac722dce7787cce0a0ded7305705.pdf. (accessed: 11.15.2019).

[5] Petra Isenberg et al. *An Exploratory Study of Visual Information Analysis.* URL: https://innovis.cpsc.ucalgary.ca/innovis/uploads/Publications/Publications/Isenberg_2008_AES.pdf. (accessed: 11.15.2019).

[6] Alation Inc. *The Alation Product.* URL: https://www.alation.com/product/. (accessed: 11.15.2019).