Time Is on My Side: A Case Study of Analyst Time Saving during Onboarding 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 saves time and effort at data-driven companies through faster onboarding of their analytics teams. We quantify the extent to which an organization has achieved some of these benefits through a specific case study of consolidation of query tools. For analysts writing SQL queries, we examine time saved by leveraging the "Compose" query tool in Alation that enables them to connect to multiple databases in a single query environment without requiring tool installation or configuration. The time savings and benefits of using the platform extend well beyond this area but are not covered in the present work.

Ninety percent of organizations using Alation leverage the Compose tool to author and execute queries, and this case study highlights the time savings that one organization achieved. Using a sample of anonymized usage data, we quantify the time users have saved due to the elimination of tool installations and connection configurations. In this organization, which has 2,423 users building queries against 61 data sources in Alation, we find a savings of more than 90 work days in one year.

Keywords: Analyst, On-boarding, Querying Databases

1. Background

How a new employee is assimilated into the organization correlates with an organization's turnover rate, profitability, product quality, etc. [1]. An effective, well-designed onboarding process enables a new employee to jump-start their learning process with minimal effort.

Organizations employ different data sources across their data landscape, and part of an analyst's onboarding process is navigating through the process of identifying, installing, and configuring the right tools to query data from these different data sources.

This case study investigates one organization's time savings and efficiency gains in the analysis process through shortened onboarding time in Alation – a software platform which (among other things) enables analyst to kick-start their query journey with the "Compose" tool which requires minimal tool installation. In organizations using the Alation platform, we find that almost 90% of organizations[1] have users writing queries using Compose.

2. Theory

Alation enables organizations to centralize their querying and database connection infrastructure, which results in users spending minimal effort setting up query tools for each specific database that their organization uses. Without Alation, each user would need to install and configure a query tool, and establish a connection to the database they need to access. In Alation, an admin performs a one-time setup process for each database connection and administers access to users. Each user can then open the Compose tool in a browser window, select the source they would like to access from a list and start querying[2]. The platform thus enables analysts to query different databases using a single environment, helping them get a holistic view of data from different sources. Alation simplifies the database management as well, since the administrator can configure access for every user; any data source migration or update to the data source connection can be handled by the admin once, without any action required by the users.

The goal of this case study is to measure the time savings through the specific area of obviating tool installation when compared to the scenario where Alation does not exist and users would need to install different query tools.

Alation also shortens learning and onboarding time in other ways that are not covered in our analysis – such as by surfacing a rich catalog of data documentation to analysts as they write their first queries. Analysts can also discover the relevant data sources by viewing others' queries or reading usage notes, without needing to seek out an expert in real-time. Significant productivity improvements have been seen (as measured by query output rate) in analysts working on Alation as a result of its collaborative nature [2].

3. Data

We use anonymized data logs summarizing when queries are executed and what types of databases users run queries on. Individual users and data objects cannot be identified through these logs. In this case study we explore how 2,423 users in a large e-commerce company, where employees write a lot of queries, save time from minimal tool installation. We look at 12 months of data containing users' first interactions with a data source to determine the annual impact of shortened onboarding time.

4. Method

We split the analysis into two scenarios – the time it would take a user to start querying without Alation and the time it takes a user to start querying on Alation.

[1] Organizations reporting usage data to Alation

[2] A desktop tool for Compose is offered for flexibility but not required

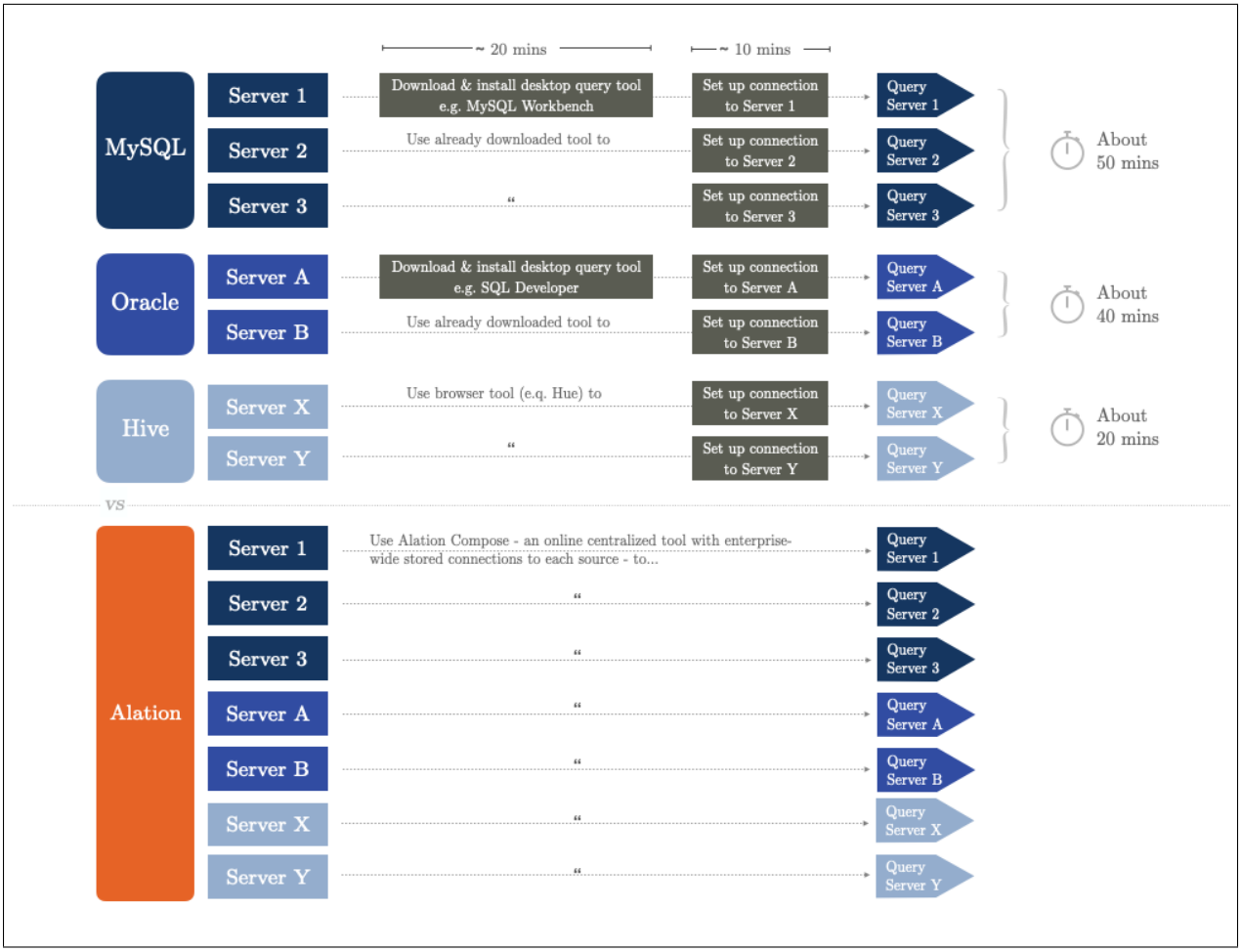


Fig. 1: Connecting to data sources without vs with Alation

4.1. Without Alation

Even though the premise of this section is hypothetical, we have actual data from customer instances for the different data sources and databases a user connects to on Alation. We use this to estimate the time it would take to install query tools for those data sources.

Table 1 shows the different databases that Alation is connected to at this organization and the number of data source connections that exist under each type. More than half the connections (32) are to different instances of MySQL databases.

Some databases, like MySQL, require a desktop tool to be installed in order to query the data and others have cloud-based query tools. From the installation documentation of common desktop query tools, it is estimated that it takes approximately 30 minutes in total to download and install [3][4] a desktop query tool and then to configure a connection to the database [5][6] that the users wants to query. This includes installing the necessary drivers, debugging and finding the right connection details

Database type	Number of data sources
MySQL	32
Oracle	11
Custom connection	4
Teradata	4
Postgres	3
Presto Hive	2
Redshift	2
SQL Server	1
BigQuery	1
Hive	1
Total	**61**

Table 1: Database types and number of connections at this organization

from internal resources. To set up any additional connections to a database (for example, if a user has already installed a query tool for MySQL to connect to a data source and needs to set up a new connection to a different MySQL instance), 10 minutes is estimated. This process involves finding and configuring the uniform resource identifier (URI), port, and credentials for the connection. Setting up the connection to a cloud-based tool is similar and hence, is also estimated to be 10 minutes. Fig. 1 illustrates a fictional scenario of a user who needs to connect to seven different data sources. Three of these connections are to different instances of MySQL databases, requiring an initial desktop installation time of 30 minutes with two additional 10 minutes required to set up additional connections. On the other hand, the Hive database has a cloud-based query tool and only needs a 10 minute configuration time.

Depending on how many types of databases and the number of instances that a user connects to, the configuration time may vary. The total tool configuration time across the organization T is calculated as follows:

$$T = \sum_{u=1}^{U} (\sum_{d=1}^{D_u} T_d + \sum_{c=1}^{C_u} T_c) \quad (1)$$

where U = number of users
D_u = number of database types for user u
C_u = number of additional database connections for user u
T_d = time for user to install query tool for database type d
T_c = time to configure a connection to a database (10 mins)

4.2. With Alation

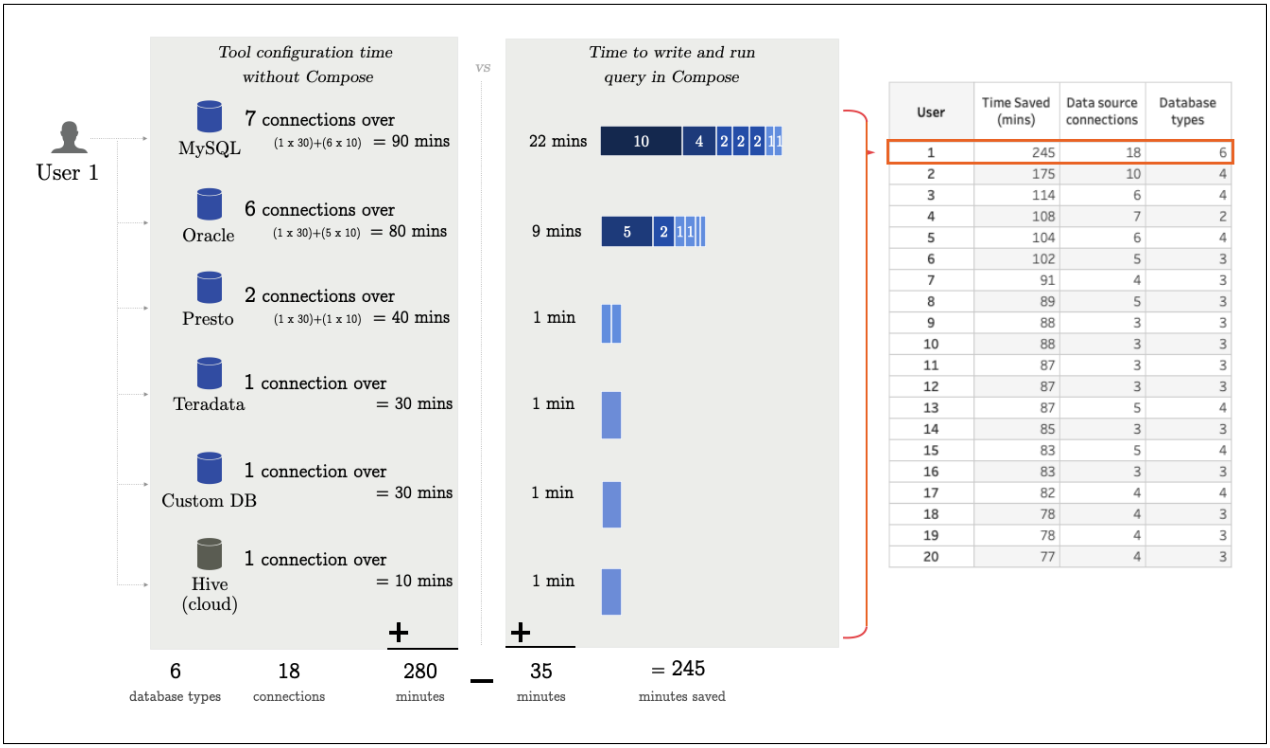
For this section, data from users' first interaction with Alation's Compose tool was observed to measure the time taken to execute their first query against a data source. That time is measured as the time between the first time they started a query against a new data source to the first time they successfully executed that query.

From the data of users entering Compose to write a query, defined by looking at the instances of users entering, writing and executing a query within the same work day (8 hours), we find that:

- 72% of users compose and run their first query against a data source in under 5 minutes
- 45% of users run their first query against a data source in under 1 minute

5. Results

Using total installation time calculated from Eqn. 1 and users' initial query time in Alation from section 4.2, we aggregate the time saved across our dataset of 2,423 users in this organization. We find that the organization saved 92 work days (734 hours) from avoiding installation of query tools across its user base. Fig. 2 shows the top 20 users that saved the most time from avoiding query tool installation and



User	Time Saved (mins)	Data source connections	Database types
1	245	18	6
2	175	10	4
3	114	6	4
4	108	7	2
5	104	6	4
6	102	5	3
7	91	4	3
8	89	5	3
9	88	3	3
10	88	3	3
11	87	3	3
12	87	3	3
13	87	5	4
14	85	3	3
15	83	5	4
16	83	3	3
17	82	4	4
18	78	4	3
19	78	4	3
20	77	4	3

Fig. 2: Top 20 time-saving users at the organization and breakdown of top user's savings

a break down of how the top user saved time. User 1 for example saved 4.1 hours (245 minutes) from connecting to 18 different data sources across 6 different database types. User 4 on the other hand, connected to 2 databases across 7 different connections, saving 108 minutes.

6. Summary and Conclusions

The Alation platform enables analysts to onboard and get to the data faster requiring no installation or configuration of query tools. In this case study, we see that:

- Alation's Compose query platform facilitates analysts' access to different data sources through a single platform, and in general users run a query within minutes of starting up the tool.
- The organization in this case study leveraged Alation's Compose tool to save 92 work days across 2,423 users across the platform over the course of a year.

Additionally, when infrastructure changes are made within companies, users querying using other desktop query tools would need to spend time reconfiguring their connections to keep up with the changes. Within Alation however, the administrator can make these configuration changes in bulk with no effort on the part of the individual user. These savings aren't accounted for in this case study.

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 [7].

References

[1] SHRM.org. *Onboarding a new employee*. URL: https://www.shrm.org/resourcesandtools/tools-and-samples/toolkits/pages/onboardingandassimilationprocess.aspx. (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] Oracle Corporation. *Installing Oracle SQL Developer*. URL: https://docs.oracle.com/cd/E39885_01/doc.40/e38928/install.htm. (accessed: 2.18.2020).

[4] MySQL. *Installing MySQL Workbench*. URL: https://dev.mysql.com/doc/workbench/en/wb-installing.html. (accessed: 2.18.2020).

[5] Oracle Corporation. *Creating a Database Connection Using SQL Developer*. URL: https://docs.oracle.com/database/121/ADMQS/GUID-0DE0C9F4-8800-4142-A755-14B2FAA6624F.htm. (accessed: 2.18.2020).

[6] MySQL. *Creating A New MySQL Connection*. URL: https://dev.mysql.com/doc/workbench/en/wb-getting-started-tutorial-create-connection.html. (accessed: 2.18.2020).

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