



THE CATALOG IS THE PLATFORM

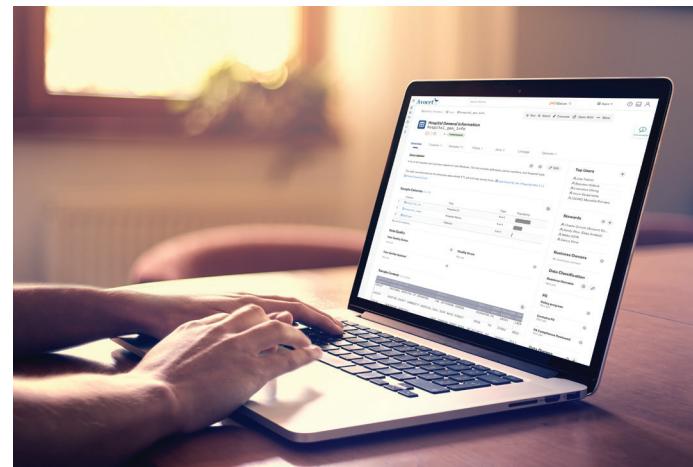
Supplying trustworthy and reliable data
to support BI and AI Initiatives

Business intelligence (BI) and artificial intelligence (AI) are important capabilities for data-driven enterprises. BI has been a staple of data-driven decision making for decades and AI is quickly becoming central to an effective data strategy. But both BI and AI depend on an emerging category of software to be successful: data intelligence. [According to market researcher IDC](#), data intelligence ensures that BI and AI are fueled by “trustworthy and reliable data.”

Without trustworthy and reliable data, BI and AI initiatives stall, struggle to create value, and can even damage the company’s reputation and bottom-line. BI has been plagued by the garbage-in, garbage-out problem since its inception. BI’s saving grace is the human element involved. If a dashboard has inaccuracies or draws incorrect conclusions, it falls to a human to intervene — obviously not ideal, but manageable on a small scale. AI, on the other hand, lacks the human element. If the data used to feed AI isn’t trustworthy and reliable, there is no human safety net to pick up the pieces — and that mistake gets replicated over and over again, causing irreparable damage.

For data intelligence to be implemented effectively, it must be available to everyone who works with data, be easy to use for a wide range of data users, and be consistent across users, so everyone is applying and sharing the same intelligence across data. Unfortunately, current approaches to data intelligence don’t make this easy. Enterprises struggle to wrangle dozens of tools into a comprehensive approach. Challenges arise on many levels: getting disparate solutions to work together and getting everyone to use them is tough, and making the business case for all solutions can be just as hard. Few people are willing to buy a car one piece at a time with the hope that someday it will drive.

The data catalog has evolved to overcome the challenges of implementing data intelligence. A data catalog is an intuitive place for everyone to work with data, making it an ideal place to implement data intelligence consistently across the organization. And, by leveraging the data catalog as a platform for data intelligence, an organization can consolidate myriad solutions — including data governance, cloud transformation, and privacy, risk & compliance

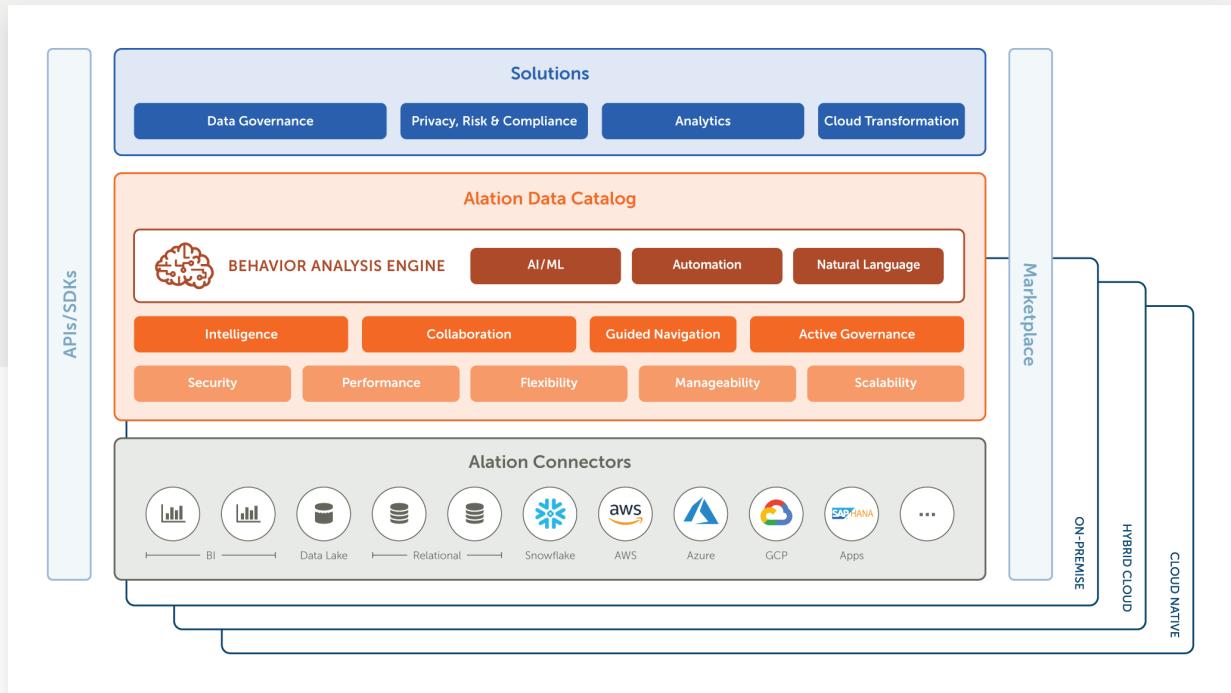


— into one place. This allows for the elimination of silos that hamstring data intelligence efforts. This also makes it easy for data workers to make the business case for a platform for data intelligence rather than a patchwork of solutions.

Today, Alation is leading the evolution of the data catalog from a single use-case application for data search & discovery to a platform that supports a broad range of data intelligence solutions.

This white paper explores capabilities that are critical to making a data catalog a platform for data intelligence, and examines how Alation supports each. In particular, these capabilities include:

- Broad Functionality to Support Data Intelligence Use Cases
 - Data Governance
 - Cloud Transformation
 - Privacy, Risk & Compliance
- Intelligence to Automatically Surface Context and Remove Manual Effort
- Enterprise-Grade Capabilities
 - Security
 - Monitoring and Management
- Modern Architecture



Broad Functionality to Support Data Intelligence Use Cases

For the data catalog to be an effective platform for data intelligence, it must support a growing set of data intelligence use cases. Let's explore three of the most pertinent use cases for modern enterprises: data governance, cloud transformation, and privacy, risk & compliance.

Data Governance

With data governance, organizations seek to ensure analytics are built atop trusted, high-quality data while reducing risk and ensuring the compliant use of data.

A data catalog can help address data governance, providing visibility, a central repository for operational and regulatory policies, and a business glossary. Greater visibility into data quality empowers data consumers to understand whether the data can and should be used. As a central repository for operational and regulatory policies, the data catalog helps consumers to understand pertinent policies and apply them to the analytics at hand. Finally,

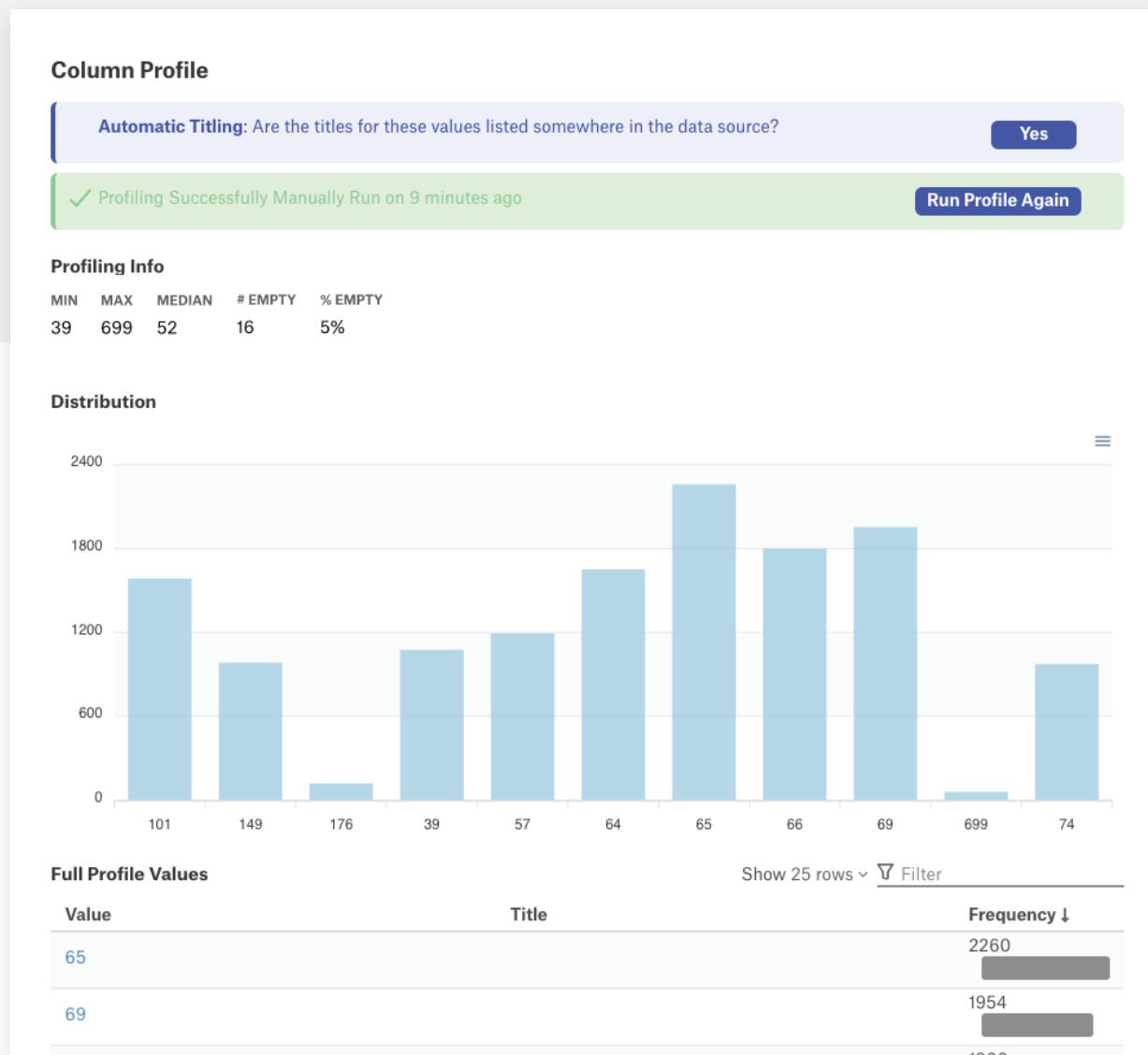
a business glossary ensures that terms are understood and used consistently across the organization. In order to address data governance, these data governance capabilities must be available in the platform.

Alation captures organizational and regulatory policies and directly associates them with the underlying data. Alation makes it easy to centrally manage data



handling requirements and also helps teach data consumers to understand how data should and should not be used. Alation's business glossary also gives users a comprehensive view of the organization's business terms. This establishes a common semantic foundation and enables the enterprise to drive accurate and consistent decisions.

Alation goes a step further than the average data catalog by supporting trusted, high-quality data in two more ways. First, Alation can profile statistics including minimum, maximum, average, and number of NULLs at the column-level for select data sets. These statistics give analysts and other data consumers a quick view into the quality of data before it's used. Second, Alation enables subject matter experts (SMEs) to apply quality flags like endorsements and deprecations to data sets. These flags are automatically surfaced at the point of data use, ensuring that data governance is actively applied all the way through the use of the data.



The screenshot shows a data catalog interface with a warning message overlay. The title bar reads "Diagnosis by State". The main area shows a SQL query:

```
1 SELECT drg, provider_state, total_discharges FROM ipps.  
  'NH', 'ME', 'VT' /* Northeast */
```

The warning message is titled "Use with Caution" and contains the following text:

⚠ Warning
This data is being moved to [Snowflake] Supply Chain and will be deprecated on September 15. Please talk to Jackson Louis (Sr. Field Engineer) if you have questions.

Mitch Perkins (Aug 18 at 2:59 pm)
This schema contains PII information.
Please reference the PII policy located here: [PII Compliance](#)

Please contact [Governance and Compliance](#) for any question
Erin Lloyd (Oct 10 at 7:39 am)

Cloud Transformation

Organizations are moving to the cloud for a variety of reasons, from reducing costs by reducing their on-premises footprint to taking advantage of new innovations only available in the cloud and leveraging this more flexible environment to support continuous integration (CI) and continuous delivery (CD).

Whatever the motivation, cloud transformation initiatives don't always go smoothly. These initiatives often incur unanticipated costs, particularly when an organization attempts a complete "lift-and-shift" strategy, when in actuality only a subset of the organization's data is actually used and worth moving. And during the migration process, organizations find themselves in a state of flux. Chaos can occur when data is migrated and spread across on-premises and cloud environments, even if the end game is a hybrid environment or 100 percent cloud. This part of the data migration process can be confusing for analysts who are no longer sure where to go for the best, most appropriate data.

A platform that supports cloud transformation must do two things: help IT prioritize which data to move to the cloud, and help data consumers find the data they need wherever it may be in the migration journey.

Alation automatically gathers usage insights, such as how often data is used, which data is most used, and which data assets are related and most relevant. IT can leverage these insights to ensure that only active, relevant data assets are moved to the cloud, reducing costs and optimizing the cloud data environment. Alation also offers a single view to help data consumers find, understand, and trust data no matter where it resides. Through advanced search capabilities and quality flags, analysts are directed to the best data, whether that data sits on-premises or in the cloud. Additionally, Alation's lineage and impact analysis give data consumers a clear view of the data as it moves and how that movement affects processes and analytical applications. Together, these capabilities help mitigate the risk and confusion associated with cloud data migration.



≡ Avocet

Search Alation

Alation

Apps ▾

Star Watch Compose Open With More

Declined Loans dclnd_loans

1 Warning 3 Endorsements

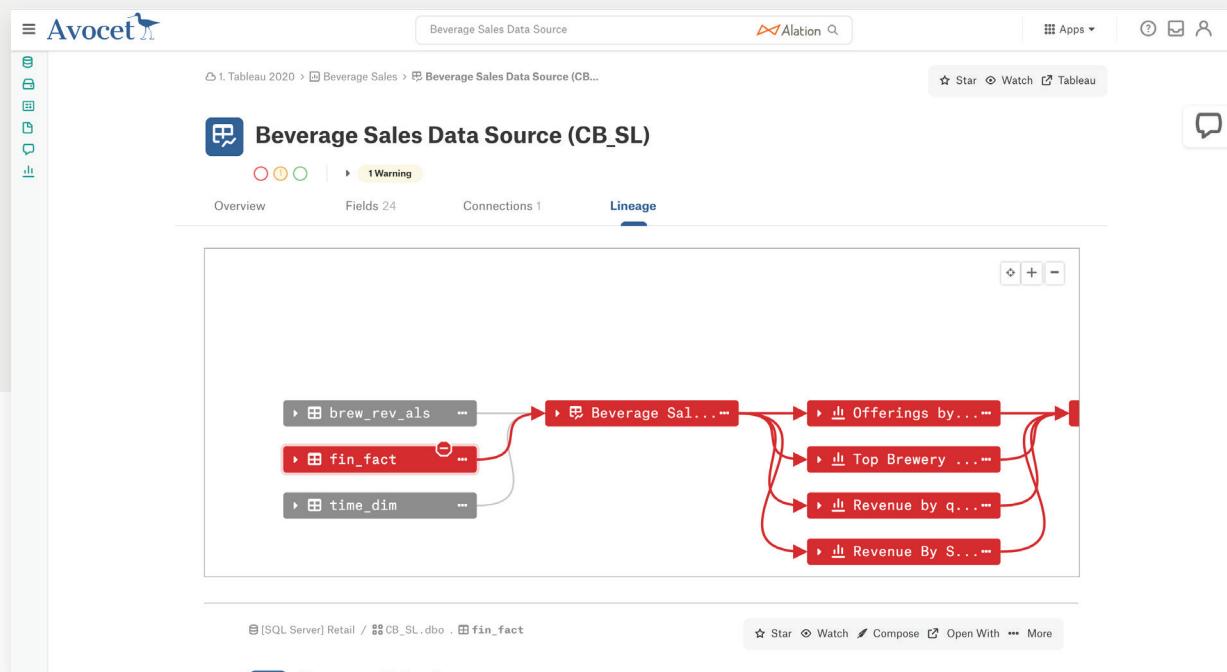
Overview Columns 10 Samples 100 Filters 2 Joins 3 Lineage Queries 2

Columns

#	★	✓	Column	Title	Type	Popularity ↓
5	★		loan_title	Loan Type	varchar(255)	
3	★		amount_requested	Amount Requested	decimal(30,10)	
6	★		risk_score	Risk Score	decimal(30,10)	
4	★		application_date	Application Date	date	
1	★		loan_id	Loan ID	int(11) No Nulls	
2	★		member_id	Member ID	decimal(40,0) No Nulls	
8	★		employment_length	Employment Length	varchar(255)	
10	★		dti_decimal	Debt To Income Ratio Decimal	decimal(30,10)	
9	★		policy_code	Policy Code	int(11)	
7	★		debt_to_income_ratio	Debt To Income Ratio	varchar(255)	

Showing 1 to 10 of 10

Page 1 Prev Next



Downstream Impact Analysis

Show objects that are impacted downstream from provider_info

Object Type is	Steward	Source is	Temp Tables are	Max Distance is	
Where All Types and Anyone and Any and Included and 5				1	
<input type="button" value="Up"/>	<input type="button" value="Down"/>				
Flag	Downstream Object	Type	Within	Steward(s)	Distance
✓	Store Inventory Kobalt	Table	[MySQL] Analytics / public.ippps	Leslie Knope	1
✓	Indiana Parks parks_213	Table	[MySQL] Analytics / public.ippps	Leslie Knope	1
⚠	Store Inventory	Report	Tableau / Distributor Revenue S...	Ron Swanson	1
✗	Medical Infrastructure Spending parks_223531_231_infra_2141323_pre_med_i...	Table	[MySQL] Analytics / public.ippps	Pam Beesley +1 more	2
✗	Store Inventory	Dashboard	Tableau / Distributor Revenue S...	Ron Swanson	2
✗	Transformation Station	Dataflow	[MySQL] Analytics	Tim Apple	2
✓	Store Name store_name	Column	[MySQL] / public.ippps.koba...	None	2
✓	st_Id	Column	[MySQL] / public.ippps.koba...	None	3
✗	Transformation Amalgamation	Dataflow	[MySQL] Analytics	Transformation...	3
✗	Indiana Parks	Dashboard	Tableau / Distributor Revenue S...	Michael Scott	3

Done

Privacy, Risk & Compliance

Privacy, risk & compliance is a wide-ranging use case that is influenced by dozens if not hundreds of regulation-specific requirements. The end goal, however, is clear: Organizations need to improve how they collect, use, and manage personally identifiable information (PII) in order to comply with regulations like GDPR and CCPA and other privacy policies, including policies created from within the organization.

To effectively manage PII, organizations must provide visibility into which data is sensitive and go a step further to instruct analysts on how that sensitive data must be used and managed. To be an effective platform, a data catalog must be able to effectively address privacy, risk & compliance, not only identifying sensitive data but guiding compliant data use as well.

Alation empowers organizations to effectively manage and protect sensitive personal data across the enterprise. Privacy and compliance policies are centrally stored within Alation, and these policies can be directly applied to the data and surfaced to the analysts at the point of data consumption. In addition, Alation can be used to discover, classify, and tag sensitive data assets at scale. And on top of the single view of sensitive data, Alation's analytics capabilities enable organizations to gather insights into usage patterns in order to quickly identify potential data misuse and gauge compliance.

The screenshot shows the Alation Data Catalog interface. The main title is "Social Security Number (SSN)". Below it, there are tabs for "Overview & Rules", "Shared Fields", and "Sync Children". The "Overview & Rules" tab is selected. The "Properties" section on the right lists the following details:

Rule Type	Conditional
Object type	Column
Created By	Raj K.
Created	8 minutes ago

The "Membership Rules" section indicates that the catalog set contains 11 columns, including attributes where column names exactly match "SSN" and exclude attributes where column names match regex "repssn".

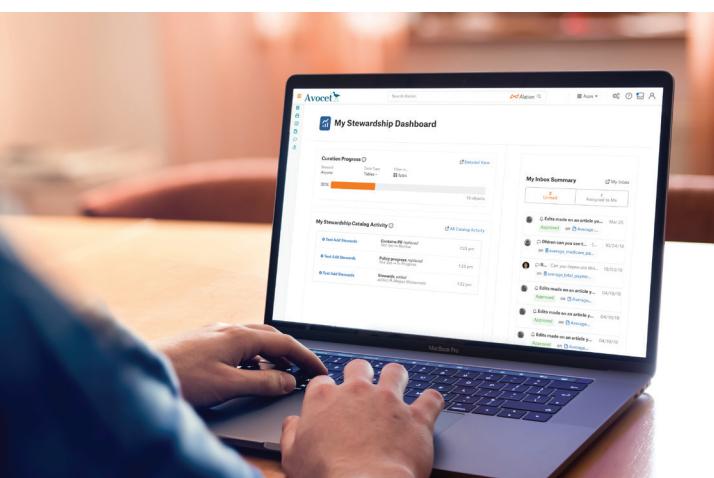
The "Catalog Set Members" section shows two entries:

Member Name	Table	Schema	Data Source
ssn	employee_acq	npi	[MySQL] Analytics
sen	applications	ferratum_nend	[AuroraDRL - Transnformation]

Intelligence

Initially, data catalogs emerged as a way to enable more people to search and discover data, but as the use cases grow more sophisticated, intelligence becomes core to a data catalog's ability to support a wide range of data intelligence use cases.

Alation recognized early that intelligence had to be applied to data search & discovery to surface relevant results. And now, Alation applies intelligence to every facet of the data catalog from data governance to cloud migration.



For example, Alation applies intelligence to data stewardship. Stewards are responsible for curating and describing data so that others in the organization can effectively understand and use it. Stewards often find it difficult to know where to focus their energy as the mountain of data under their purview always outstrips their available bandwidth. Alation examines

query logs to understand how data is being used. By combining insights into the popularity of data sets along with their level of curation, Alation shows data stewards exactly where they should focus their efforts.

"Alation gives us the visibility to contend with the enormous amount of data our business generates, prioritizing our stewardship efforts and applying data policies at scale to all of our data sets. With Alation, we have a single source of reference to monitor which business units have made progress or not in improving data sets, giving us greater visibility into department-level stewardship efforts and measure the impact these efforts have on our analytics."

Markus Schmidberger
Director of Data Technology at Scout24

The screenshot shows the Alation My Stewardship Dashboard. On the left is a sidebar with icons for Home, Overview, Catalog, Data Quality, and Help. The main area has a header "My Stewardship Dashboard".

- Curation Progress:** Shows a progress bar for "Steward A, Stew Ward" at 57%. It includes filters for "Data Type All types" and "Filter in... Select Filter". A "Detailed View" link is available.
- My Stewardship Catalog Activity:** A table listing catalog activity:

Event	Description	Time
[MySQL] Analytics / 55 ips...	Policy progress replaced	2:45 pm
[MySQL] Analytics / 55 ips...	Business Domains added	2:45 pm
[MySQL] Analytics / 55 ips...	Description replaced	2:45 pm
[MySQL] Analytics / 55 ips...	Stewards added	Sep 4
[MySQL] Analytics / 55 ips...	Stewards removed	Sep 4
[MySQL] Analytics / 55 ips...	Stewards removed	Sep 4
- My Inbox Summary:** Shows 3 unread messages and 2 assigned to me. The messages are:
 - Doe... If so, and you plan on... on Customer Demograp... 11:35 am
 - Can we augme... I need ag... on dem_growh_r... Sep 4
 - Doe... I have an analysis th... on strm_anlytcs_dm (Str... Sep 4

Another example of how Alation applies intelligence is business glossaries. Alation's business glossary provides a single location for business semantics, enabling cross-departmental access to business-approved definitions. Building and maintaining glossaries is challenging, especially given the sheer number of terms and names that exist within most organizations. Alation applies intelligence to this problem and simplifies the creation of glossaries by auto-suggesting terms based on the analysis of the names and abbreviations found in schemas, tables, and columns. There are dozens of more examples of how Alation applies intelligence to complex use cases to augment the efforts of those who work with data.

Enterprise-Grade Security, Monitoring, and Management

When new systems are introduced into the environment, IT administrators must consider both the security implications and the overhead associated with ongoing maintenance. Alation is built from the ground up to meet enterprise-grade needs around security, monitoring, and management.

Security

Introducing any new system incurs risk. Data catalogs are no exception. While data catalogs may not hold the underlying data itself, they do hold metadata, including both technical metadata (schema information) and, in the case of Alation, behavioral metadata (information about how the data is being used). When used for data governance or privacy, the data catalog houses quality flags and sensitivity classifications, which are often closely associated with underlying data. For these reasons, data catalog security must be built-in, not bolted on.

Deep investments in product, organizational, and operational security make Alation the data catalog of choice for some of the most security-sensitive enterprises around the globe. Alation's strong product security is underpinned by a strong authentication model, where users can access Alation through a customers' existing directory and identity providers. Single sign-on is supported through LDAP- and SAML-based authentication. Authorization is addressed at a granular level. Data catalog content can be restricted at the table level, ensuring that users only see catalog pages and search results to which they have access. From a data protection standpoint, Alation uses industry-standard encryption with AES256 and TLS, which can be limited to TLS v1.2.



For organizational and operational security practices, Alation is certified against ISO 27001. This ensures appropriate practices are in place for the secure development, storage, and transmission of data needed for the development of Alation, including incident response and periodic penetration testing.

Monitoring and Management

As the data catalog is used by more people across the organization for a wider range of use cases, it must be highly available and easy to monitor and manage. The data catalog must plug into existing enterprise systems,

eliminating one-off or redundant costs related to areas like security and uptime.

Monitoring the health of Alation is made simple with out-of-box capabilities, which empower administrators to easily check the status of catalog-related services. Through a monitoring dashboard, for example, administrators can check message queues related to metadata extraction (MDE) jobs and even throttle lower priority tasks to ensure the completion of higher priority jobs. Administrators are also given real-time visibility into who is using the catalog, minimizing disruption if services need to be brought down.

Started	Status	Detail
Jan 23 2020 at 10:37am	SUCCEEDED	Completed metadata extraction from Healthcare, extracted npi, ipps, Pharmaceuticals schemas. Took 1.34 seconds.
Jan 23 2020 at 10:37am	SUCCEEDED	Completed metadata extraction from Healthcare, extracted npi, ipps, Pharmaceuticals schemas. Took 1.33 seconds.
Dec 10 2019 at 10:10pm	SUCCEEDED	Completed metadata extraction from Healthcare, extracted ipps, npi, opioid, Pharmaceuticals schemas. Took 2.71 seconds.
Dec 9 2019 at 10:37am	SUCCEEDED	Completed metadata extraction from Healthcare, extracted npi, ipps, opioid schemas. Took 2.99 seconds.

Alation makes managing the data catalog straightforward. According to research conducted by Forrester,* Alation requires less than one FTE IT administrator to manage the data catalog. Alation gives administrators the ability to view and take action on the data catalog using role-based access

*Forrester: The Total Economic Impact™ Of The Alation Data Catalog, 2019

control (RBAC) and fine-grained permissions. The catalog can be organized based on different roles, including viewers, stewards, composers, and multiple types of administrators. Administrators are able to enforce policies by role, including who can view certain catalog pages and who can modify certain fields.

"According to research conducted by Forrester, Alation requires less than one FTE IT administrator to manage the data catalog."

Just as important, Alation seamlessly plugs into most enterprises' IT environments, integrating with directory and identity providers. With enterprise user provisioning, administrators can pre-configure mappings between LDAP groups to Alation roles, automating user management through groups in customers' enterprise directory. Finally, in the event of a failure in the primary environment, a separate standby or high-availability server can help relaunch catalog services.

Modern Architecture

According to IDC, by 2022 more than 90% of enterprises worldwide will use a mix of on-premises/dedicated private clouds, multiple public clouds, and legacy platforms to meet their infrastructure needs.** Most organizations have de facto hybrid environments. With that in mind, organizations should be able to deploy the data catalog wherever works best for them, whether in the cloud, on-premises, or some combination.

"According to IDC, by 2022 more than 90% of enterprises worldwide will use a mix of on-premises/dedicated private clouds, multiple public clouds, and legacy platforms to meet their infrastructure needs."

At the same time, the data catalog should scale with increased demand. As the data catalog is used by more types of users, including non-technical

**IDC MarketScape: Worldwide Data Catalog Software 2020 Vendor Assessment, 2020

business users, the system gets taxed more heavily — more articles, more conversations, and potentially more data sources. The system must remain performant, ideally in a way that does not involve installing more instances or upgrading the server. In short, the data catalog must boast a modern architecture that allows for both deployment flexibility and scalability.

Alation can be deployed on-premises or in any cloud. Customers can deploy in whichever environment works best according to their organizational needs, such as regulatory constraints or cost considerations. Alation also boasts a cloud-native architecture, helping drive performance by decoupling

services and letting customers spin up additional containers as needed. Another way Alation ensures high performance is by allowing tasks to be handled asynchronously. For example, queries can be scheduled through Alation to run during off-peak hours.

Alation is built to meet even the most taxing performance requirements from some of the



largest global organizations. One Alation customer has more than 4,000 users and 30 connected data sources. Another customer has run more than six billion queries over six million columns in 500,000 tables.

Extensibility Through Open and Programmable Interfaces

When the data catalog is the platform for data intelligence, it must allow for extensibility and customization through open and programmable interfaces. Open interfaces allow customers, partners, and even the vendor to easily extend the platform. The result is technology that can be tailored to meet company- or use-case-specific needs. Additionally, open interfaces allow for integration with best-of-breed solutions from partners and the ability to take advantage of innovations from the broader marketplace.

Alation promotes extensibility through open and programmable interfaces that build on top of the data catalog's capabilities and increase its reach across the enterprise, including a wide array of native connectors, which are available out-of-box. Using pre-built connectors reduces time to value and makes it easier to begin immediately collecting metadata from commonly used data sources. Alation also provides an Open Connector Framework, enabling third parties, including customers and partners, to build connectors to additional data sources in cases where connectors are not already available.

Additionally, open APIs supercharge many of Alation's core capabilities. For example, these APIs can be used to programmatically apply and update data quality indicators based on customizable rules. These APIs can also exchange metadata with lineage tools like Manta, allowing for advanced, cross-system lineage.

While Alation is often the starting point to find and understand data, it is not the only tool analysts use. Alation can expose certified data sets to tools like Tableau via TrustCheck. Analysts that begin their journey with Alation can seamlessly continue working with the data they find in tools like Dataiku.

The screenshot shows the Alation Data Catalog interface. On the left is a sidebar with navigation links: Home, Explore (which is selected), Favorites, Recents, Recommendations, Users, Groups, Schedules, Jobs, and Tasks. The main content area is titled 'Explore / Sales / Avocet Data Set'. It displays the 'Avocet Data Set' details, including the owner 'AlationSales', a green checkmark indicating it is 'Certified by AlationSales', and a note that it was published by 'Avocet Data Set'. A 'LIVE CONNECTION' button is highlighted with a red box. Below this, there are sections for 'Ask Data', 'Connections 1', and 'Connected Workbooks 1'. The 'Ask Data' section shows a search bar and a list of fields: 'Abc ABV', 'Abc Account', 'Abc Account Category', '# Acct ID', '# Acct ID (Accounts ...)', 'Abc Data ID', 'Abc Data ID (Location)', '# Data ID (Quarter)', '# Data ID (Region)', '# Data ID (State)', and '# Date ID (Fiscal Peri...'. To the right, there is a 'Questions To Ask' section with a 'Learn More' link, a 'Basic Data Analysis' section with a 'sum of Gross Profit' and 'by City' option, and a box containing the text 'Ask about fields in this data source'.

The Alation Data Catalog Is the Platform

Alation pioneered the modern data catalog market and is leading the evolution of the data catalog to become a platform for data intelligence. No single tool can serve all of an enterprise's data intelligence needs. Instead, enterprises need a platform that supports data intelligence.

According to IDC, "Rather than collecting all the diverse, distributed, and large-scale data into one common platform, intelligence about the data becomes the platform."** Not only is Alation the best data catalog for overcoming the challenges of modern data management and creating a data culture, but also the best platform for data intelligence use cases.

From data governance to privacy, risk & compliance, and cloud migration, Alation applies intelligence throughout the process of working with data. Alation offers enterprise-grade security, monitoring, and management and is built on a modern cloud-native architecture that enables deployment flexibility, scalability, and performance — all in fully extensible, open, and programmable interfaces. The data catalog is the platform for data intelligence, and Alation is the leading data catalog as a platform.

[**IDC MarketScape: Worldwide Data Catalog Software 2020 Vendor Assessment, 2020](#)