Ali Daneshmand

Ali.daneshmand@SAICcorp.com

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.   
When you’re ready to add your content, just click here and start typing.]

Alation Best Practices

FOR SAIC

Table of Contents

[INTRODUCTION 3](#_Toc82374949)

[Change Log 3](#_Toc82374950)

[I.T. / System Administrator Definitions & Best Practices 4](#_Toc82374951)

[I.T. Best Practices 4](#_Toc82374952)

[What Are Deployment Tiers? 4](#_Toc82374953)

[Properties of SAIC’s Alation Deployment Tiers 5](#_Toc82374954)

[Standard S.A. vs Alation S.A. Definitions 6](#_Toc82374955)

[System Analyst / Administrator (S.A.) Services 6](#_Toc82374956)

[S.A. Software Installation 6](#_Toc82374957)

[S.A. Monitoring 7](#_Toc82374958)

[S.A. Configuration Management 8](#_Toc82374959)

[S.A. User Management 8](#_Toc82374960)

[S.A. Networking 8](#_Toc82374961)

[S.A. System Improvement 9](#_Toc82374962)

[S.A. Miscellaneous Standard Responsibilities 9](#_Toc82374963)

[Upgrade Alation 10](#_Toc82374964)

[Quick Start – Alation Overview for New Users 12](#_Toc82374965)

[What Are Data Catalogs? 12](#_Toc82374966)

[Why Alation Data Catalog? 12](#_Toc82374967)

[General Usage Theme – Steward or Analyst 13](#_Toc82374968)

[Data Sources & Access Rights 13](#_Toc82374969)

[Compose & Queries 13](#_Toc82374970)

[Articles, Glossaries & Taxonomies 13](#_Toc82374971)

[Alation Feature “How-To’s” 14](#_Toc82374972)

[Data Sources 14](#_Toc82374973)

[ How To Connect to Data Source(s) 14](#_Toc82374974)

[ QLI Ingestion 14](#_Toc82374975)

[ MetaData Extraction 14](#_Toc82374976)

[ Data Source Access Rights via Alation 14](#_Toc82374977)

[Queries 15](#_Toc82374978)

[ Compose 101 15](#_Toc82374979)

[ Queries 101 15](#_Toc82374980)

[Articles & Glossaries 16](#_Toc82374981)

[Article vs Glossary 16](#_Toc82374982)

[Taxonomy 16](#_Toc82374983)

[Templates 16](#_Toc82374984)

[Alation Feature Based Best Practices 16](#_Toc82374985)

[Using Trust Flags: Proceed With Confidence 16](#_Toc82374986)

[Best Practice #1: Determine who should have the ability to flag 17](#_Toc82374987)

[Best Practice #2: Stewards with admin privileges should monitor the trust flags 17](#_Toc82374988)

[Best Practice #3: Provide a rationale when “warning” or “deprecating” a data object 17](#_Toc82374989)

[Best Practice #4: Display alternative data sources 17](#_Toc82374990)

[Best Practice #5: Consider when to use endorsements 18](#_Toc82374991)

[Best Practice #6: Encourage users to flag filters, joins, and queries 18](#_Toc82374992)

[Curating Catalog Pages: Clarify Your Data 18](#_Toc82374993)

[Best Practice #1 - Prioritize catalog pages to curate 19](#_Toc82374994)

[Best Practice #2 - Create custom templates for each type of catalog page 19](#_Toc82374995)

[Best Practice #3 - Start with a small number of “gold standard” catalog pages 20](#_Toc82374996)

[Best Practice #4 - Assign stewards to curate top priority catalog pages 20](#_Toc82374997)

[Best Practice #5 - Set and re-visit goals 20](#_Toc82374998)

[Working With Articles 21](#_Toc82374999)

[Best Practice #1: Consider information needs from the users’ perspective 21](#_Toc82375000)

[Best Practice #2: Create templates for articles that belong together 22](#_Toc82375001)

[Best Practice #3: Summarize articles in glossaries for easy reference 23](#_Toc82375002)

[Best Practice #4: Make it easier to find articles using taxonomy 24](#_Toc82375003)

[Best Practice #5: Consolidate existing content from disparate sources 25](#_Toc82375004)

[Best Practice #6: Create “gold standard” articles as a model for others 25](#_Toc82375005)

[Best Practice #7: Determine who can and should author articles 25](#_Toc82375006)

[Best Practice #8: Link articles to other parts of the data catalog 25](#_Toc82375007)

[Best Practice #9: Avoid creating redundant articles 26](#_Toc82375008)

[Access, Roles, & Permissions in Alation 27](#_Toc82375009)

[Proper Instructions on Access Management 27](#_Toc82375010)

[Roles 28](#_Toc82375011)

[Hats 29](#_Toc82375012)

[List of Suggested Custom Roles & Responsibilities 29](#_Toc82375013)

[Access to Objects 31](#_Toc82375014)

[Visibility Settings Via Manual Catalog Sets 32](#_Toc82375015)

[Access to Catalog Fields 32](#_Toc82375016)

[Access to Data 33](#_Toc82375017)

[Per-Object Parameters 33](#_Toc82375018)

[Sensitive Data Setting 34](#_Toc82375019)

[Dynamic Profiling 35](#_Toc82375020)

[Obfuscate Literals 35](#_Toc82375021)

[Access to Compose 36](#_Toc82375022)

[Back-Up & Restore 37](#_Toc82375023)

[1. Virtual Machine(s) 37](#_Toc82375024)

[2. On-Premise Machine(s) 37](#_Toc82375025)

[General Query Management 37](#_Toc82375026)

[Understanding Alation Analytics: Object Queries 37](#_Toc82375027)

[Metadata Extraction 38](#_Toc82375028)

[APPENDIX 38](#_Toc82375029)

[REFERENCES 38](#_Toc82375030)

[Alation Usage 38](#_Toc82375031)

[System Administration / Management Best 39](#_Toc82375032)

[B.O.K. Best Practices 39](#_Toc82375033)

[GLOSSARY 39](#_Toc82375034)

# INTRODUCTION

### Change Log

|  |  |  |  |
| --- | --- | --- | --- |
| **Index** | **Change Date** | **Person Who Made Change** | **Change Made** |
| 1 | September 12, 2021 | Ali Daneshmand | Original Draft Document |
| 3 | … |  |  |

This document covers Best Practices as it pertains the usage of Alation Data Catalog at SAIC Corp. LLC.

# I.T. / System Administrator Definitions & Best Practices

By definition, System Manager or System Administrator is a person who is responsible for the upkeep, configuration, and reliable operation of “computer systems”; especially *multi-user computers*, such as “servers”. The system administrator seeks to ensure that the “uptime”, “performance”, “resources”, and “security” of the computers they manage meet the needs of the “users”, without exceeding a set *budget* when doing so. In the case of Alation Data Catalog and SAIC, a textbook System Administrator is exactly who is needed for set up, maintenance, upgrades, end-user needs taxonomy / needs, etc.

Below are two general a comparison of a generic set of services typical System Administrators perform, versus the technical needs of Alation Data Catalog.

## I.T. Best Practices

### What Are Deployment Tiers?

Generally, in Software Engineering and in Content Management Systems (to which Alation Data Catalog can technically fall under each category), based on User Feedback and Software Architects, the software in question undergoes ongoing improvement in the form of releases. This requires “Deployment Tiers” in both the Engineering Company AND the client company, but from different perspectives. From the perspective of the “client” (in this case SAIC), Development Tiers is a forum in which the newest releases are tested for robustness of existing and new features, bug fixes from previous versions, robustness of current content on the new software, etc. The exact layout and procedures are dictated based on the needs of the specific client. Some examples of “Development Tiers” might look like this …:

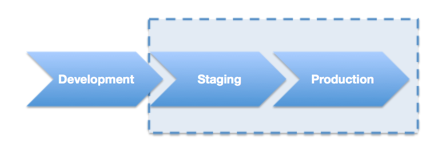


Figure 1: Example of a "Development Tier"

A picture containing card, drawing

Description automatically generated

Figure 2: Another Example of a "Development Tier"

### Properties of SAIC’s Alation Deployment Tiers

|  |  |  |
| --- | --- | --- |
| **#** | **SAIC’s Alation Development Environment** | **SAIC’s Alation Production Environment** |
| 1 | Stand-Alone Instance | High Availability (Active / Passive) |
| 2 | Standard Installation | Standard High Availability Installation |
| 3 | Connect to All Applicable Data Sources for testing purposes ONLY! | Only connect the Data Sources that will be used for day-to-day work. |
| 4 | Day-To-Day Work: No day-to-day work should be conducted here by Data Scientists / Data Analysts / Business Analysts. This environment is dedicated to I.T. work and for End-User Feature Testing on new and existing features. This is where IT can test upgrades, test new Data Source Connections, troubleshoot issues raised in production, etc. | Day-To-Day Work: No feature testing work should be run here. This environment is dedicated to the Data Scientists / Data Analysts / Business Analysts who run queries, develop Metadata Models, run metadata Extractions, manage all data, manage all reports, etc. |
| 5 |  | By nature of Alation’s “High Availability” structure, this “Production Environment” also doubles as a “Disaster Recovery” environment. Please refer to the instruction manual for details on Alation’s “High Availability”. |

## Standard S.A. vs Alation S.A. Definitions

### System Analyst / Administrator (S.A.) Services

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Services)** | **Alation System Administrator Tasks (Services)** |
| 1 | Back Up Protocols | 1. Standard S.A. Back Up Protocols 2. Alation Specific Back Up Methods |
| 2 | Mail Services | Knowledge of Mail Services Protocols available for the Data Catalog. |
| 3 | NFS | 1. Setting Up NFS isn’t necessary. But knowledge of how NFS, and a Network Contact with NFS Admin access is needed here. 2. Past clients have used NFS based “back-up drives” and “data drives”. |
| 4 | Web Services (httpd service) | 1. Knowledge of Web Services required 2. Knowledge of SSL / HTTPS / SSL Certificates |
| 5 | DNS Protocols | General DNS Knowledge would be helpful. |
| 6 | Database Administration | 1. General Data Knowledge is needed 2. Network Table Knowledge for Connectivity purposes needed. |

### S.A. Software Installation

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Software Installation)** | **Alation System Administrator Tasks (Software Installation)** |
| 1 | Application Installation | Standard Application Installation |
| 2 | O.S. Installation | Standard O.S. Installation |
| 3 | User Customization | Standard User Customization |
| 4 | Software Packaging | Standard Software Packaging |

### S.A. Monitoring

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Monitoring)** | **Alation System Administrator Tasks (Monitoring)** |
| 1 | System Monitoring | * Alation provides API’s that allow for Alation level Monitoring * Alation SA should also implement standard SA/SE-BOK System Monitoring Procedures, expecially for Memory Usage, RAM Usage and CPU Utilization |
| 2 | Resource Accounting | Alation provides a definition of the minimum “Resources” it requires. The SAIC Alation System Administrator should be aware of these resources at time of the Data Catalog Instllation. |
| 3 | Data Display |  |
| 4 | Network Monitoring | This should be performed by default by existing SAIC / ComEd / PECO / etc “Network Administrators” and “Network Administrators” |
| 5 | Benchmark Testing | This is a useful Testing Practice to allow appropriate meaning for future System Performance and Metrics. |
| 6 | Configuration Discovery | Configure discovery methods to find resources to manage from your network, Active Directory, and Azure Active Directory (Azure AD). First enable and then configure each method that you want to use to search your environment. You can also disable a method by using the same procedure that you use to enable it. |
| 7 | Host Monitoring |  |

### S.A. Configuration Management

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Configuration Management)** | **Alation System Administrator Tasks (Configuration Management)** |
| 1 | Site Configuration | Alation provides detailed instructions for each release of its Data Catalog. |
| 2 | Host Configuration |
| 3 | Site Move |  |
| 4 | Fault Tolerance | * Alation Data Catalog Software Already has some robust Fault Tolerance. But it’s Fault Tolerance is mainly dependent on it’s Hardware Profile. * SA should regularly monitor system to ensure that the system resources do not deteriorate. |

### S.A. User Management

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (User Management)** | **Alation System Administrator Tasks (User Management)** |
| 1 | Accounts | Along with the Alation SME, and the Alation Governance Committee, the Alation System Administrator should define an Access Management Model and a criterion of granting / managing access to current and new users |
| 2 | Documentation | * This is self-explanatory. Documentation should be Standard Operating Procedures. * In this case, Documentation can proverbially “Piggyback” onto Alation. * This documentation can take many forms, ranging from Attribute discussions, to Steward Ownership to Articles and everything in-between. |
| 3 | Policy |  |
| 4 | User Interaction |  |
| 5 | White Pages |  |

### S.A. Networking

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Networking)** | **Alation System Administrator Tasks (Networking)** |
| 1 | Network Configuration |  |
| 2 | LAN |  |
| 3 | WAN |  |
| 4 | Host Tables |  |

### S.A. System Improvement

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Improvement)** | **Alation System Administrator Tasks (Improvement)** |
| 1 | Models |  |
| 2 | Software Design |  |
| 3 | Training Administrators |  |

### S.A. Miscellaneous Standard Responsibilities

|  |  |  |
| --- | --- | --- |
| **Index** | **Generic System Administrator Tasks (Miscellaneous)** | **Alation System Administrator Tasks (Miscellaneous)** |
| 1 | Trouble Tickets |  |
| 2 | Secure Root Access |  |
| 3 | General Tool |  |
| 4 | Security |  |
| 5 | File Synchronization |  |
| 6 | Remote Access |  |
| 7 | File Migration |  |
| 8 | Resource Cleanup |  |
| 10 | Configuring, adding, and deleting file systems. |  |
| 11 | Ensuring parity between dev, test and production environments. |  |
| 12 | Training users |  |
| 13 | Plan and manage the machine room environment |  |

### Upgrade Alation

#### V R5 To V R6

Update from V R5 (5.9.x) to V R6 (5.10.x) follows the [R](https://alationhelp.zendesk.com/hc/en-us/articles/360036573373-Updating-Alation-to-V-R6-5-10-x-#h_809676f2-1cfb-4d30-9347-42f8aa99310b)egular Update Scenario.

#### Regular Update Scenario

Please plan your update from V R2, V R3, V R4, and V R5 using the following information:

* General Alation update information: Updating Alation
* Important Update Safety Best Practice
* Update instructions for your instance type:
  + Manual Software Update - Standalone Systems
  + Update on HA Pair (4.7 and Above)
* PostgreSQL upgrade from 9.3 to 9.6: Upgrading Internal PostgreSQL Instances From 9.3 To 9.6.

#### Irregular Scenario: MongoDB Migration is Not Complete

Incomplete off-Mongo DB migration causes the following error during the attempted update to V R6:

The installer will abort until migration job from MongoDB to Postgres has succeeded.

If this happens in your instance, contact [Alation Support.](mailto:support@alation.com) Alation Support will help to find the root cause for the Mongo migration being incomplete and to move forward with the update.

#### PostgreSQL Upgrade to 9.6

Applies to instances UPDATED to V R6 from prior releases

Does not apply to new installations of V R6

Once you have updated Alation to V R6 you can choose to upgrade its internal PostgreSQL databases to version 9.6. This PostgreSQL upgrade can be done on a schedule independent from updating the Alation application.

|  |
| --- |
| **Note**  Releases prior to V R6 use PostgreSQL 9.3.  PostreSQL 9.3 is at end of life and no longer receives updates or bug fixes. |

We recommend that you upgrade to PostgreSQL 9.6. Alation will stop using the PostgreSQL 9.3 in one of the releases that will follow V R6 and upgrading PostgreSQL to 9.6 will become a requirement.

#### Scheduled Query Migration: V R6 Changes

There are some important changes to the behavior of scheduled queries in V R6 compared to previous releases. Please be prepared for the following:

#### R4 and earlier

##### **If you have UNPUBLISHED SCHEDULED queries**

In R6, you can no longer schedule unpublished queries.

During update to V R6, your UNPUBLISHED SCHEDULED queries will be automatically PUBLISHED in order to keep your query schedule. You will receive an email notification informing you about this change and containing the list of queries which have been auto published.

Note that this auto-publishing will create a major version in the query version history.

##### **If you have PUBLISHED SCHEDULED queries that include changes you haven’t published yet**

In R6, only the published part of the query can run on schedule. During update to V R6, such queries will be automatically republished in full so that there are no unpublished edits in these queries. You will receive an email notification informing you about this change and containing the list of queries which have been auto republished.

Note that auto-republishing will create a major version in the query version history.

#### Related Articles

**Updating on HA Pair with Splitting the Cluster** - [*https://alationhelp.zendesk.com/hc/en-us/articles/360025479673*](https://alationhelp.zendesk.com/hc/en-us/articles/360025479673)

# Quick Start – Alation Overview for New Users

This section is a general overview of Data Catalogs and Alation specifically. This section is meant for those who have “heard” of Alation and Data Catalogs and have never used the tool. Other sections will go into further details on similar topics.

The next major section is an “How-To” section. The

## What Are Data Catalogs?

By definition, a “data catalog" belongs to a database instance and is comprised of metadata containing database object definitions like base tables, synonyms, views or synonyms and indexes. The SQL standard lays down a regular method for accessing the data catalog known as the information schema, though not all databases use this. They may implement other features of the SQL standard. A data catalog ensures capabilities that enable any users, from analysts to data scientists or developers, to discover and consume data sources. It includes a crowdsourcing arrangement/model of metadata and annotations that permits every user to contribute their knowledge.[]

## Why Alation Data Catalog?

Alation is a data catalog that provides a sophisticated set of Metadata Management Functionality that has been acknowledged by Gartner as well as a sophisticated set of social features for a high level of End-User collaboration and Data Self-Service. These features include but are not limited to …:

* Built In Agile Approval Process for Articles
* Built In Agile Approval Process for Glossaries
* Focused Conversations functionality regarding Articles, Glossaries, Queries, Data Objects, Data Sources, etc.
* Tagging Alation Users, Alation Groups, and other Alation based Objects in Articles and comments.
* Advanced Content Sharing between registered users

## General Usage Theme – Steward or Analyst

## Data Sources & Access Rights

Access to [data sources](https://alationhelp.zendesk.com/hc/en-us/articles/360038060834-Types-of-Sources-in-the-Alation-Catalog) (databases of various types) can be controlled for each individual data source and granted to individual users. Whether or not you can view a data source and its metadata in Alation will depend on its visibility settings.

A data source can be **public** (visible to everyone) or **private** (visible only to allowed users).

To be able to work with the settings of a data source and run the data jobs you must be granted the **Data Source Admin** function for a given data source. A user with any Alation role can be made a Data Source Admin.

Please see [Access Tab](https://alationhelp.zendesk.com/hc/en-us/articles/360038675133-Access-Tab) for more details about access settings of data sources.

## Compose & Queries

## Articles, Glossaries & Taxonomies

# Alation Feature “How-To’s”

## Data Sources

### How To Connect to Data Source(s)

### QLI Ingestion

### MetaData Extraction

### Data Source Access Rights via Alation

## Queries

### Compose 101

### Queries 101

## Articles & Glossaries

### Article vs Glossary

### Taxonomy

### Templates

# Alation Feature Based Best Practices

This section shows “Best Practices” for some common Alation Features.

## Using Trust Flags: Proceed With Confidence

*Trust Flags* (also known as the Red/Yellow/Green stoplights) are the primary way in which trust of data can be communicated to users. Trust Flags are available at the data source (including file systems and BI reports), schema, table and column levels, as well as for filters, joins, and queries.

Users can click on the green stoplight to *endorse* a data object. They can *warn* other users about potential issues with data by clicking the yellow light and adding a required explanation. Similarly, users can *deprecate* a data object by clicking the red light and explaining why the data object is not to be trusted.

In addition to highlighting critical data issues on the catalog pages, trust flags are now carried forward to SQL writing in Compose via the *Trust Check* feature. SQL writers will see highlighted text if data they are using for their queries has been endorsed, warned or deprecated. Analysts can quickly and easily trust that the data that they are using is appropriate and proceed with confidence.

Some of the ways we have seen customers use trust flags:

* Endorse "gold standard" queries
* Warn about data sets with non-continuous date ranges
* Deprecate data sets that are being decommissioned

### Best Practice #1: Determine who should have the ability to flag

For most organizations we recommend a bottom-up approach: enabling the majority of users to endorse, warn, or deprecate data objects. When users are properly trained, this maximizes the collective knowledge of the user community. For example, if a user identifies a problem with data that they know best they can take action and issue a warning or deprecation so that the community is immediately informed.

Some organizations may believe that only stewards have the combination of knowledge and authority required to assess whether a particular data object should be trusted. Some organizations may elect a top-down approach where only stewards are permitted to edit trust flags. Every element that can be flagged (except for filters, joins, and queries) can have permissions attached to restrict the ability to edit to the designated individuals or groups. The downside to this approach is that if permissions are restricted then the stoplights are not visible, and, as a result, users will not have as prominent a visual cue for trust flags. To supplement the knowledge of stewards, all users should be encouraged to be on the lookout for data quality issues and if they see something, they should say something by communicating with the appropriate steward through the Alation conversation feature.

### Best Practice #2: Stewards with admin privileges should monitor the trust flags

Having knowledgeable stewards monitor trust flags combines the best of the top-down and bottom-up approaches of data curation. The knowledge of the community is captured, but the risk of uncontrolled trust flags is mitigated. If there is disagreement among users about whether data should be trusted or not, the stewards with admin capability should intervene to clarify the discussion and drive to a decision. If necessary, the steward can delete trust flags that are inaccurate or obsolete.

### Best Practice #3: Provide a rationale when “warning” or “deprecating” a data object

When a data element is warned or deprecated the user will be prompted to complete a required text field. This is where the rationale for the warning or deprecation should be documented. For example, if there is a problem with the data pipeline, it should be indicated in the text field for the warning or deprecation. In addition to clearly stating the problem, best practice would be to include appropriate contacts, links to articles, and time frames for resolution or data source decommissioning (if applicable).

Warnings should be used where data is essentially accurate, but certain precautions should be taken or conditions accounted for. For example, if there’s a data set that has accurate data but only over a certain time frame, that should be noted in a warning. On the other hand, deprecations should indicate data that is fundamentally flawed and should not be used, at least for the time being.

### Best Practice #4: Display alternative data sources

Often when one data source is deprecated, it is because there is another, better source that should be used. By using a link to the preferred data set in the deprecation, Alation can be used to point users to the data that is trustworthy.

*Endorse - Warn – Deprecate*

A screenshot of a cell phone

Description automatically generated

### Best Practice #5: Consider when to use endorsements

Warning and deprecating data are clearly to be used when there are exceptions and data need to be flagged as untrustworthy for one reason or another. Depending on the current state of your data, stewards may consider using endorsements in a similar matter. For instance, if a data source has many instances of *untrustworthy* data, it may be necessary to proactively endorse data that is good. In this case *trustworthy* data can be considered the exception that needs to be called out via the endorsement trust flag.

Another time to consider using endorsements is where there are two similar data sets that may be "in competition". In this case endorsements can prevent potential misunderstandings about which data set to use. This is a situation that shouldn’t last for long as the "losing" data set should be deprecated and, ultimately, decommissioned.

### Best Practice #6: Encourage users to flag filters, joins, and queries

All users have the ability to flag filters, joins, and queries. Users have the responsibility to be on the lookout for problems with data and how it is used in queries and should use the flags to document their concerns. As an example, if there is a poorly constructed query that is consuming inordinate DB resources, it should be flagged, with a link to a "gold standard" model query included in the warning explanation. Just as with data catalog pages, stewards with admin privileges should monitor and control trust flags so they are not used inappropriately.

## Curating Catalog Pages: Clarify Your Data

Catalog pages hold information about a company’s data objects, including its data sources, schemas, tables, columns, directories, files, queries, and BI content. For each data object, the catalog page includes title, description, and fields for additional information. All catalog pages, except for queries, have templates that can be customized with custom fields.

Catalog pages are the primary way that Alation conveys specific data object information to its users. With Alation, information about data is documented in one place, reducing the need to rely on tribal knowledge for users to gain an understanding of their data sources.

Some of the ways we have seen customers curate catalog pages:

* Documenting queries so new or inexperienced analysts can understand and re-run existing queries.
* Curating data sets so that key resources are not asked the same question repeatedly.

### Best Practice #1 - Prioritize catalog pages to curate

Once data sources are connected and metadata extracted, there may be many tens of thousands of catalog pages. Companies should prioritize the catalog pages based on business criticality, complexity, and usage.

Business criticality can be assessed based on the importance of the data set to the business. This is likely to be data sets that meet high-level corporate objectives, such as revenue enhancement, cost savings, or regulatory compliance.

Another driver for prioritization is complexity. For instance, if a data set generates an outsized number of questions, it should be considered for priority curation so that critical resources do not need to answer the same questions repeatedly.

Another driver of prioritization is how often a data set is used. Alation’s Query Log Ingestion will measure the relative use of schemas, tables, and columns. Alation presents the results in the "popularity" bar for each data object. High usage catalog pages are a natural place to focus efforts.

***Object Popularity***

### Best Practice #2 - Create custom templates for each type of catalog page

Carefully consider the information requirements for each type of catalog page. The information contained on catalog pages serves two purposes:

1. helping to find data elements
2. defining the data elements once found.

To help users find data elements, think about how the typical user would search for what he or she is looking for. As an example, if users are searching based on certain geographies then there should be a custom field for region so that when users do advanced searches they can quickly narrow their search. Other custom fields should be added that will aid either in both free text search and advanced search.

Once users have found the catalog page for the data element they are looking for, the custom template should include fields that provide all of the information that a typical user would need. For example, if users are concerned with data that is personally identifiable, there could be a PII field that has a multi-picker indicating "PII" or "Non-PII" or "Unclassified". As another example, we have a pharmaceutical company that has a custom field for "molecule" so that its analysts know that they are working with the data set relevant for their drug investigations.

To create the custom template, a core team of knowledgeable users should convene in workshops and come to agreement on the custom fields that should be included for each type of catalog page for both searching and data clarification.

### Best Practice #3 - Start with a small number of “gold standard” catalog pages

Once the templates are defined, the core team should set out to create a small number of catalog pages for each type that will serve as the "gold standard" or model for other stewards to follow when curating a catalog page.

Virtually all catalog pages will include a description of the data object. Descriptions should be complete and to the point. Including descriptive commentary for data objects will enhance search results and enable self-service. The descriptions should leverage the power of Alation by making links to other parts of the catalog, including articles, related catalog pages, as well as external links, where appropriate. Additionally, links should be included to subject matter experts who have knowledge that is complementary to the assigned stewards.

For schemas, tables, and columns, Alation may attempt to assign natural language names based on the machine language names captured during metadata extraction. For the gold standard catalog pages the titles should either be confirmed or entered.

Finally, all built-in and custom fields should be completed.

***Sample Curated Catalog Page***

### Best Practice #4 - Assign stewards to curate top priority catalog pages

Stewards should be selected based on their knowledge of the data sets covered by the catalog pages. Good candidates for stewards may be Top Users of the data set, which Alation identifies during Query Log Ingestion. Email the proposed stewards and explain their responsibilities and include links to the gold standard catalog page.

***Top Users***

### Best Practice #5 - Set and re-visit goals

The goals for curation of catalog pages will vary considerably by organization, driven by the number of data objects to be cataloged as well as the capacity of the people on the team. Whatever the organization’s particular goals, the core team should communicate the goals for the extended team as well as goals for the individual stewards.

One approach to goal setting could be based on a segmentation of the prioritized data objects. For instance, an organization could set as its goals to document 100% of critical items, 80% of important items, and 20% of the other data objects. You should breakdown these organization-wide goals to the individual curators who will be doing the documentation.

Whatever the organization’s goals are initially, they are likely to change over time and should be re-visited periodically. As new data sources are added, you will need to add these to the mix, prioritize, and assign to stewards.

## Working With Articles

Articles are Alation’s method to house information that is broadly relevant for Alation users and a complement to catalog pages which hold information specific to a particular data object. Articles have free text fields that contain ample space to describe a topic thoroughly, as well as custom fields that add discrete pieces of information, such as people, dates, or any other sort of classification. These custom fields provide additional content for articles while also helping to narrow searches for relevant articles.

Articles are important because they are a permanent, accessible, and searchable store of information about a company’s data. Unlike other legacy stores of information, articles are in the context of data, so users don’t have to switch applications to get the information they need. Also, the articles can be linked to data objects they are relevant to so that they are always current.

To learn more about how to create articles in Alation, see [Working With Articles](https://alationhelp.zendesk.com/hc/en-us/articles/360012402333-Working-with-Articles).

Some of the ways we have seen customers use articles:

* Presenting the company’s data strategy
* Documenting governance policies, such as how to handle PII
* Defining KPIs and key business terms
* On-boarding topics and tutorials
* Presenting detailed discussions of results of analysis, including links to data sets, queries, and dashboards

### Best Practice #1: Consider information needs from the users’ perspective

Alation curators should understand and segment their user base and create content required for each segment. Some example user segments could include data scientists, analysts, or data stewards. Other segments could be experienced or novice analysts, or, perhaps, analysts in various functional areas, such as sales, marketing, or supply chain.

For each segment the Alation curator should understand what information a typical user needs to do their work. For instance, one pharmaceutical company’s analysts needed to understand milestones for clinical trials by region, which they defined in articles.

Users often make their information needs known by asking questions of key resources either via messaging software or by face-to-face meetings. These questions that have general interest are good topics for documenting in articles so time can be saved for both questioner and answerer through self-service. Plus, unlike with traditional methods, answers documented in articles will be available for use in the future.

As Alation instances are populated, curators should understand any content gaps that exist for each user segment and fill them with relevant articles that will live in Alation. As curators create articles in Alation they should consider these questions:

* What is the purpose of the article you are creating?
* In what way will it help users understand a particular topic?
* What is the most relevant and important information that users need to know about a particular subject? (Keep it concise!)
* Think about how you want to organize articles so they are clear and easy to understand.

### Best Practice #2: Create templates for articles that belong together

Templates contain customizable fields that make groups of articles easier to find and their content easier to understand. Curators should group articles that logically belong together and create templates for each group. Examples of groupings of topics might include data policies, on-boarding tutorials, or KPIs for each geographic region or product line. For each grouping, the curator should work with stakeholders to define the template for the grouping.

The custom fields in templates can make articles easier to find in search and help focus results in advanced search. For example, each on-boarding article could have a field indicating its functional area, so analysts in marketing, sales, or supply chain could enter the search term to find tutorials relevant to them.

Once the article is found, the custom fields of the template will contain information that is relevant to the topic being described in the article. For instance, the on-boarding articles may all have 30-60-90 days plans, which can be fully described using a free-text field. KPIs, on the other hand, may have fields for term definition, approvals, and region codes.

***Custom Templates and Custom Fields***

A screenshot of a cell phone

Description automatically generated

### Best Practice #3: Summarize articles in glossaries for easy reference

Alation has the feature for displaying selected fields for articles that share a template in a Glossary form. Curators should consider which information would be useful for users to see summarized in a glossary. The type of information that is suitable for a glossary is usually where there are many members of a group and a high-level summary is useful. For example, an organization may have hundreds of acronyms related to data. Users could reference a glossary to understand what acronyms stand for and then click on the underlying article if they wanted more details.

***A Glossary***

A screenshot of a cell phone

Description automatically generated

### Best Practice #4: Make it easier to find articles using taxonomy

As users explore information, they will usually start at a high-level and drill deeper as they learn and look for more detailed information. Alation supports grouping topics that have a parent-child relationship so that if a topic logically falls under another topic, it can be designated as the "child". After the relationship is defined, child articles will be shown underneath parent articles in the Taxonomy view, facilitating guided navigation. They are very flexible, allowing for any number of sub-categories.

***Article Taxonomy***

A screenshot of a cell phone

Description automatically generated

### Best Practice #5: Consolidate existing content from disparate sources

Before they deployed Alation, the majority of our customers had multiple sources of information about their data. One pharmaceutical company said that before Alation, their environment was like a "cesspool" with important information being kept in multiple, disparate sources, including Google Docs, Sharepoint websites, and Powerpoint presentations. With information in so many sources, users are likely to miss critical data or give up in frustration.

Admins should survey all existing sources of information and consolidate relevant and accurate articles in Alation. Curators should be asking the same questions as with new content when this information is brought into Alation (see [Best Practice #1](https://alationhelp.zendesk.com/hc/en-us/articles/360017554593-Writing-Articles-Retain-Tribal-Knowledge#h_775014270151549499544291)).

### Best Practice #6: Create “gold standard” articles as a model for others

After curators have decided on their target audience and useful topics, it is time to begin writing articles that will serve as models for the rest of the data community. We have some customers who have employed technical writers to begin the process of article creation. Although this is an excellent practice, most organizations will not have such resources available. However, typically there are members of the core team that are eager to share content they have created previously or write new articles that serve a clear need. These users should create "gold standard" articles that employ best practices described here. The particulars of what constitutes a "gold standard" will vary by industry and organization, but the goals of knowledge documentation and collaboration are similar everywhere.

### Best Practice #7: Determine who can and should author articles

After an initial set of gold standard articles are published, additional authors should be recruited. Determining who can write articles will be dependent on the organization’s view of data democratization. At Alation, we recommend keeping access to data as open as possible to encourage contribution and collaboration among the entire data community. In our view, this holds for both article authorship and data object curation. We would encourage all members of the community to write articles on topics that they know particularly well and that generate a lot of questions. They should follow the guidelines in this document to ensure that articles are necessary, properly structured, included in the appropriate templates, and not redundant.

As a control on article authorship, curators may consider including a form of workflow into articles. They could include a custom field that would indicate that the article was reviewed and approved. The field itself can be access protected. With this approach the knowledge of the entire community can be collected while maintaining quality standards.

### Best Practice #8: Link articles to other parts of the data catalog

One of the most powerful features of articles is the ability to embed any other object that is cataloged in Alation using @-mention. For example, users can see an expandable, live preview of the data object that the article is referencing. Additionally, when users go directly to that data object they can see articles that are linked to it, which they can access to get additional information about the data object they are viewing.

***@-Mentioning Alation Objects***

A screenshot of a cell phone

Description automatically generated

### Best Practice #9: Avoid creating redundant articles

As existing documentation is brought into Alation, admins should be careful to eliminate any redundant content.

* SEARCH the catalog to see what articles already exist.
* CONFIRM the content you find aligns with your understanding of the topic.
* UPDATE article content as needed to ensure accuracy.
* EXTEND article content (e.g. add more detail, create children articles) to provide deeper understanding of the topic.

#### Case Study

A large eCommerce company was rapidly expanding its analytics teams. On-boarding these new analysts was an ad-hoc process, with junior analysts shadowing their more experienced colleagues. This was a time-consuming and inefficient process for both the senior and junior analysts and the results were inconsistent.

With Alation articles, the company captures the knowledge of its experienced analysts and presents it in a structured and consistent format. The articles contain links to other articles that are relevant to analysts, for example, policies for data management and security. Embedded in each article are links to data sets that are important for new analysts to understand and start out with. New analysts can also see links to other analysts within the organization who are willing to serve as mentors and discuss specific issues. These conversations can be maintained on Alation so that future analyst classes will have a record for their reference.

With this approach analysts are climbing the learning curve quicker resulting in a shorter time to delivering value. With high turnover among data analysts, it’s crucial that they spend less time to gain mastery over the data landscape by learning in the same environment as their data catalog.

## Access, Roles, & Permissions in Alation

### Proper Instructions on Access Management

**NOTE: *Applies from release V R3 (5.6.x) up!***

Roles, permissions and various access rules usually define what a user is allowed or denied in a software product. In Alation, there are a number of factors that control what you can see and what you can do in your data catalog.

To control access to functionality, Alation has **roles**. To control access to **catalog objects**, there are access settings that are applied on the level of specific objects or can be inherited from the source database or BI platform. There are also **catalog field-level permissions** which define a user’s ability to edit catalog fields. Field permissions can be applied to individual users or on the group level.

A screenshot of a cell phone

Description automatically generated

### Roles

When you first join Alation, you are assigned a user **role**. You can be a Server Admin, a Catalog Admin or a Standard User:

A screenshot of a cell phone

Description automatically generated

Your user role is the primary access type that determines what pages you see in the catalog and what actions you can perform. For example, as a Server Admin, you will be able to [add sources to Alation](https://alationhelp.zendesk.com/hc/en-us/articles/360038645233-Adding-a-Data-Source), and this is something that a Catalog Admin or a Standard User will not be able to do. As a Catalog Admin, you can create and manage [catalog sets](https://alationhelp.zendesk.com/hc/en-us/articles/360038809454-Catalog-Sets) and [field permissions](https://alationhelp.zendesk.com/hc/en-us/articles/360037749074-Custom-Field-Permissions), and this is something that a Standard User will not be able to do. If an action in Alation is not allowed because your role does not grant permission for it, certain parts of the interface will appear inactive and you will see an "insufficient permissions" message on hover-over:

A screenshot of a cell phone

Description automatically generated

But no worries, Standard Users! Even without the various admin functions, you will be able to do a lot in Alation: search and find your data, create your own articles and collaborate on articles with colleagues, curate fields, participate in conversations, and author, contribute to, and run queries in Compose: most of the functionality in Alation is open to users with any role.

Please also refer to [Roles Overview](https://alationhelp.zendesk.com/hc/en-us/articles/360011997654-Roles-Overview) for a general overview of Alation user roles in releases up to V R6 (5.10.x).

### Hats

If enabled in your Alation catalog, [Hats](https://alationhelp.zendesk.com/hc/en-us/articles/227443068-Hats-in-Alation) are a way to slightly modify the layout of the main toolbar to make some features more accessible to you depending on your primary tasks in Alation. Namely, hats will make either **Compose** or the **Data Stewardship** features more accessible to you from the main toolbar. Please note that Hats do not control access to features, but they do impact the layout of the controls you see on the main toolbar. For example, here is how wearing the **Analyst** hat changes the main toolbar:



And here’s the view for a user wearing the **Curator** hat:



Note that hats do not modify your primary Alation role in any way. They only help you access either the Compose or the Stewardship features quicker from the main toolbar.

### List of Suggested Custom Roles & Responsibilities

|  |  |  |
| --- | --- | --- |
| **Index** | **Role Title** | **Responsibilities** |
| 1 | Alation System Administrator | * *Curator Hat & Alation Server Admin* * Pre-Installation Checklist * Right Start Style Installation * Initial Right Start Style Training * Virtual Server Based Back-Up & Restore Procedures and Model * System Architecture Hosting   + Knowledge of H.A. Clustering and Stand-Alone Architecture   + Knowledge of Virtual Systems, Cloud Services and On-Premise Systems * Alation System Health Checks   + Virtual Machine Monitoring   + Alation Performance Monitoring |
| 2 | Alation Subject Matter Expert | * *Curator Hat & Alation Server Admin* * User / Group Access Management * Enforce Alation Best Practices * Have Proficient Knowledge of Alation Data Catalog Software * Have Knowledge of Data Science and Data Governance * Alation-Based Back-Up & Restore Procedures and Model * Alation Performance Monitoring |
| 3 | Alation Overall Database Subject Matter Expert | * *User & Curator Hat* * Proficient Knowledge of SAIC required Data Sources * Proficient Knowledge of Alation’s Data Source Connectivity * Proficient in DB and Alation Access Management |
| 4 | Alation Data Attribute / Asset Custodian | * *User & Curator Hat* * Proficient Knowledge Of Business Rules * Proficient Understanding & Ownership of assigned Data * Knowledge of relevant Attribute / Asset Metadata * Must be responsible for QLE, Metadata Extraction, Dynamic Profiling, etc. * Collaborate with assigned Steward |
| 5 | Alation Data Attribute / Asset Steward |  |
| 6 | Alation Metadata Modeler |  |
| 7 | Data Scientist |  |
| 8 |  |  |

### Access to Objects

#### Access to Data Sources

Access to [data sources](https://alationhelp.zendesk.com/hc/en-us/articles/360038060834-Types-of-Sources-in-the-Alation-Catalog) (databases of various types) can be controlled for each individual data source and granted to individual users. Whether or not you can view a data source and its metadata in Alation will depend on its visibility settings.

A data source can be **public** (visible to everyone) or **private** (visible only to allowed users).

To be able to work with the settings of a data source and run the data jobs you must be granted the **Data Source Admin** function for a given data source. A user with any Alation role can be made a Data Source Admin.

Please see [Access Tab](https://alationhelp.zendesk.com/hc/en-us/articles/360038675133-Access-Tab) for more details about access settings of data sources.

#### Access to Articles

Articles in Alation have[access settings of their own](https://alationhelp.zendesk.com/hc/en-us/articles/360014071654-Granting-Access-to-Your-Articles-V-R4-and-later-). You can make your article private (not find-able and not viewable by other catalog users via search), you can share it with select users, or you can keep it public and accessible by everyone in Alation. In other words, your ability to find and read an article in your Alation catalog depends on the access settings of this individual article.

#### Access to Queries

Query objects have [access settings](https://alationhelp.zendesk.com/hc/en-us/articles/360036695434-Granting-Access-to-Your-Queries-in-Compose) too. Whether or not you can find a query using search and view it depends on whether or not you have access to the underlying data source. Whether or not you can **edit** a query, depends on your access rights to this individual query. Note that currently in Alation (still applies to the latest version V R6), all users can find queries for data sources they are allowed access to via search. If anyone shares a query on a data source that you do not have access to, you will see the 403 error ("access denied") when you try to view this query following the shared link.

If you do have data source access, then both published and unpublished queries on this source can be viewed in the catalog, but to be able to contribute to a query, you need to explicitly [request the Edit access](https://alationhelp.zendesk.com/hc/en-us/articles/360037266373-Requesting-Access-to-Queries) (applies to V R6 5.10.x).

#### Access to BI Objects

For BI sources in Alation (such as, for example, Tableau sources), by default Alation will attempt to establish a match between Alation users and the BI system users so that view permissions for the extracted BI objects can be inherited into Alation too. If Alation cannot find a BI user match for an Alation user, then this user will not be able to view the extracted BI objects. See, for example, [Permission Mirroring for Tableau](https://alationhelp.zendesk.com/hc/en-us/articles/360039307933-Permission-Mirroring-for-Tableau).

### Visibility Settings Via Manual Catalog Sets

Sometimes Catalog Admins may choose not to include specific data objects into search results and catalog browsing using visibility settings in [manual catalog sets](https://alationhelp.zendesk.com/hc/en-us/articles/360038811054). This is another reason why certain objects may be hidden from common use.

### Access to Catalog Fields

#### Custom Field Permissions

Whether or not users can edit catalog fields on data objects as part of the curation effort can be controlled with [custom field permissions](https://alationhelp.zendesk.com/hc/en-us/articles/360037749074-Custom-Field-Permissions). This can be part of the data governance practices in your organization defined by your Catalog Admins. If the editing capabilities on a field on a data object page appear disabled to you, this means field permissions have been enforced and editing is only allowed to a specific group of users:



### Access to Data

Alation provides an ability to view a small sample of real data within the catalog. This means catalog users may see examples of real data on the respective pages of Tables and Columns under **Samples**:

A screenshot of a social media post

Description automatically generated

There are a number of ways Data Source Admins can control access to data:

### Per-Object Parameters

It is possible to specify which data objects should NOT be browsable using the catalog browser (the left-hand navigation panel) and should NOT be sampled. To do so, you can use the settings on the **Per-Object Parameters** tab of the data source settings page:

A screenshot of a social media post

Description automatically generated

### Sensitive Data Setting

It is possible to mark data as sensitive on the column level, on the **Per-Object Parameters** tab. Values of sensitive columns will not be sampled and shown in the catalog:

A screenshot of a social media post

Description automatically generated

Sensitive columns in the **Sample Content** table on a Table object page:

A screenshot of a cell phone

Description automatically generated

### Dynamic Profiling

You can turn on **Dynamic Profiling** on the **General Settings** tab of the data source settings to require users should enter their own credentials before they can run a profile on a table and retrieve data samples. They will only be able to see the data they have access to on the database. Sample data will only be visible to the user who initiated profiling:

A close up of a logo

Description automatically generated

### Obfuscate Literals

You can turn on **Obfuscate Literals** on the **General Settings** tab of the data source settings so that literal values in ingested queries are not shown in the UI. Instead, users will see placeholder values.

Obfuscate Literals OFF:

A screenshot of a cell phone

Description automatically generated

Obfuscate Literals ON:

A screenshot of a cell phone

Description automatically generated

### Access to Compose

In Compose, before users are able to run queries on a data source and retrieve any real data, they will need to first authenticate with their database credentials. This ensures that users only retrieve data they can access on the database.

A screenshot of a cell phone

Description automatically generated

## Back-Up & Restore

The “Best Practice” here really depends on the topography of the Alation Architecture.

### Virtual Machine(s)

**Type of Back-Up / Restore:** Whether Alation is “Stand Alone” or “HA”, the best method of backing up and and restoring is through taking snapshots of the relevant virtual machines, including all relevant drives (i.e. Alation Install Drive, Alation Data Drive, Alation Back-Up Drive). Though it would be nice to take an Alation based “Back Up”, in the case of a Virtual Machine style set up, it’s not needed. This makes the restoration process much easier in that to restore Alation will mean simply restoring the Virtual Machine to the date / time where it was working properly last.

**Full Back-Up vs Partial Back-Up:** Best back up model here would be to take a Full Snapshot weekly or just before an upgrade. A “Partial” Back-Up daily to capture the detailed differences between major backups while not consuming as much disk space.

### On-Premise Machine(s)

Whether Alation is “Stand Alone” or “HA”, the best method of back

## General Query Management

* Managing Queries In Compose
* What Happens When Queries Are Published?
* How To Start A New Query?
* Working With Alation Analytics

## Understanding Alation Analytics: Object Queries

## Metadata Extraction

# APPENDIX

## REFERENCES

### Alation Usage

|  |  |  |
| --- | --- | --- |
| **#** | **Name of Reference** | **Reference URL** |
| 1 | Alation Main Page | <http://www.alation.com> |
| 2 | Alation – Query Management | <https://alationhelp.zendesk.com/hc/en-us/search?utf8=%E2%9C%93&query=Query+Management&commit=Search> |
| 3 | Alation – Best Practices | <https://alationhelp.zendesk.com/hc/en-us/search?page=1&query=Best+Practices&utf8=%E2%9C%93#results> |
| 4 | Alation – Glossary | <https://alationhelp.zendesk.com/hc/en-us/articles/360038232374-Alation-Glossary> |
| 5 | Alation - Writing Articles: Retain Tribal Knowledge | <https://alationhelp.zendesk.com/hc/en-us/articles/360017554593-Writing-Articles-Retain-Tribal-Knowledge> |
| 6 | Alation - Curating Catalog Pages: Clarify Your Data | <https://alationhelp.zendesk.com/hc/en-us/articles/360017429314-Curating-Catalog-Pages-Clarify-Your-Data> |
| 7 | Alation – API Resource Catalog | <https://alationhelp.zendesk.com/hc/en-us/articles/360011671333-API-Resource-Cataloging> |
| 8 | Alation - Updating on HA Pair with Splitting the Cluster | <https://alationhelp.zendesk.com/hc/en-us/articles/360025479673> |
| 9 | Alation – Updating Alation to R6 | <https://alationhelp.zendesk.com/hc/en-us/articles/360036573373-Updating-Alation-to-V-R6-5-10-x-> |
| 10 | Technopedia | <https://www.techopedia.com/definition/32034/data-catalog> |

### System Administration / Management Best

|  |  |  |
| --- | --- | --- |
| **#** | **Name of Reference** | **Reference URL** |
| 1 | University of Missouri (St. Louis): IT Management Department Best Practices | <http://www.umsl.edu/~lacitym/bestmis.html> |
| 2 | Pulseway System Administrator Tips | <https://www.pulseway.com/blog/5-best-practices-for-system-admins> |
| 3 | Wikipedia - Definition of System Administrator | <https://en.wikipedia.org/wiki/System_administrator> |
| 4 | Wikipedia – S.A. Duties | <https://en.wikipedia.org/wiki/System_administrator#Duties> |

### B.O.K. Best Practices

|  |  |  |
| --- | --- | --- |
| **#** | **Name of Reference** | **Reference URL** |
| 1 | DAMA Book Of Knowledge | <https://dama.org/content/body-knowledge> |
| 2 | System Administrator Book Of Knowledge |  |
| 3 | Alation Book Of Knowledge |  |
|  |  |  |

## GLOSSARY

General Glossary Terms For This Document

|  |  |  |
| --- | --- | --- |
| Index | Term | Definition |
| 1 | Data Catalog |  |
| 2 | Best Practice |  |
| 3 | System Administrator |  |
| 4 | Subject Matter Expert | a.k.a. S.M.E. |
| 5 | MetaData |  |
|  | MetaData Extraction |  |
|  | MetaData Modeler |  |
|  | Dynamic Profiling |  |
|  | Reference Document |  |
|  | Data Dictionary |  |
|  | API |  |
|  | Alation Analytics |  |
|  | Open Queries |  |
|  | Published Query |  |
|  | Unpublished Query |  |
|  | Steward |  |
|  | Deployment Tier(s) |  |

Alation Glossary Terms

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **Agile Approval** | A [collaboration workflow](https://alationhelp.zendesk.com/hc/en-us/articles/360014178213) enabling Alation users to request reviews and provide approvals for **Article** objects ensuring that articles are credible sources of information. |
| **Alation Analytics** | [Database that surfaces catalog usage data](https://alationhelp.zendesk.com/hc/en-us/sections/360003193413-Alation-Analytics) in order to provide a way for governance officers and analysts to implement better adoption of the Alation catalog and drive more value at their company from exploring ways their users interact with the data. |
| **Alation Catalog** | The component of the Alation software that serves to catalog and describe data sources offering a number of applications on top (such as Agile Approval or Stewardship Dashboard). Is often referred to as "the Catalog". Also see **Compose**. |
| **Alation Object** | An entity in the Alation catalog that can have a dedicated catalog page. Almost everything you see in Alation is an object, for example: a data source object, a table object, a query object, or an article object. |
| **Alation Search** | A set of features that allow searching and finding objects in the Catalog. |
| **Allie the Robot** | Mascot of the machine learning algorithm in Alation. Machine learning is used to suggest comprehensive titles for data objects. You can see Allie's image in various messages in Alation, and you will also see the "robot head" icon         next to an auto-suggested title. |
| **Alation Role** | When you hear "[Alation roles](https://alationhelp.zendesk.com/hc/en-us/articles/360011997654)", it is usually one of the roles a user can be assigned when they join Alation: **Standard User**, **Catalog Admin**, or **Server Admin**. Please also see **Data Source Admin**. |
| **Article** | An object in the Alation catalog that can be used for data documentation. You can create an article to store some long-form written content, such as detailed data policies or analytical insights. Articles can be added to **Business Glossaries** that help bind related pieces if information together. |
| **Built-in Field** | A field on a catalog page [template](https://alationhelp.zendesk.com/hc/en-us/articles/360038214553) that cannot be removed or modified. Usually contains property values. |
| **Catalog Admin** | A role in Alation. Granted to users whose primary task is curating the catalog and supervising the curation effort by other users. |
| **Catalog Page** | A dedicated page for an Alation object that you see in the Alation Catalog. Catalog pages are based on customizable [templates](https://alationhelp.zendesk.com/hc/en-us/articles/360038216053) that contain built-in and custom fields. Users enrich their catalog by providing field values: information describing the Alation object. |
| **Catalog Page Template** | Template used to "assemble" a catalog page from built-in and custom fields. Alation has two types of catalog page templates: object templates and article templates (or custom templates). [Article templates](https://alationhelp.zendesk.com/hc/en-us/articles/218228447) can be created by users. Both types of templates can be customized by adding custom fields. |
| **Catalog Set** | Alation object that enables you to group data objects together and perform bulk-edit operations on custom fields. Bulk edits are done on the Catalog Sets page, rather than on the data object page. |
| **Column Lineage** | Visual representation (a diagram) of how tables, views and BI reports are derived from other tables, [at the column level](https://alationhelp.zendesk.com/hc/en-us/articles/360038286653). Also see **Lineage**. |
| **Compose** | Alation's query tool. You can access Compose either using your web browser or by downloading the Alation Compose app to your computer. |
| **Conversations** | Alation enables you to publicly ask and answer questions through [Conversations](https://alationhelp.zendesk.com/hc/en-us/articles/360015070073), which live on any data object, query, result set, or user profile. |
| **Custom Field** | A catalog [field created by a user](https://alationhelp.zendesk.com/hc/en-us/articles/217377288) that can be added to catalog page templates. |
| **Custom Template** | Catalog page [template for the article object](https://alationhelp.zendesk.com/hc/en-us/articles/218228447). Can be created by Catalog Admins and modified by adding/removing custom fields. |
| **Customer Portal** | An Alation-managed site for Alation admins. Accessible with a login you receive when you install Alation at [https://customerportal.alationdata.com](https://customerportal.alationdata.com/) |
| **Data Source** | An Alation object representing a database connection in the Alation catalog. |
| **Data Source Admin** | Function at the data source level for accessing and editing the settings of a data source. [Data Source Admins](https://alationhelp.zendesk.com/hc/en-us/articles/360038675133) can view and edit the data source settings, including the connection properties and parameters for metadata extraction, profiling, and query log ingestion. |
| **Data Steward** | A field on catalog pages of data objects that stores the names of users who are stewarding this data object in the Alation Catalog. The Data Stewards field can be used for filtering the reports on My Stewardship Dashboard and the curation reports (**Curate** menu > Curation Progress, Catalog Activity, Data Objects Without Stewards). |
| **Dynamic Profiling** | When enabling dynamic profiling:   * Profiles will be user specific and user initiated. * Profiling will use user's database credentials. * Automatic profiling will be disabled |
| **Filters** | Alation learns from you every time you run a query. If you join tables or use a data expression, Alation will automatically pick it up and enable you to save it for quick reuse. On the table Catalog page, all previously used JOIN clauses are on the **Joins** tab. Data expressions are on the **Filters** tab. |
| **Glossary** | An object that is a "collection" of articles based on the same custom template. Articles [can be grouped together into a glossary](https://alationhelp.zendesk.com/hc/en-us/articles/360022917094) when they are related to one topic and are complementary to each other |
| **Hats** | A way to compile a view of the Alation interface to better suite your goals in Alation. There are [2 Hats](https://alationhelp.zendesk.com/hc/en-us/articles/227443068): Data Analysts and Data Stewards. The Analyst Hat makes the features for Analysts easily accessible from the Alation toolbar (the Compose button), and the Data Steward Hat gives access to the Stewardship features (the Curate menu). Hats work in addition to Alation Roles: a user with any Alation role can wear one or the other hat. Note that you can disable this feature. |
| **Joins** | Alation learns from you every time you run a query. If you join tables or use a data expression, Alation will automatically pick it up and enable you to save it for quick reuse. On the table Catalog page, all previously used JOIN clauses are on the **Joins** tab. Data expressions are on the **Filters** tab. |
| **Lexicon** | A system job that matches abbreviations found in the names of data objects with the comprehensive words, or expansions. Lexicon is used for auto-titling and creating [Suggested Terms](https://alationhelp.zendesk.com/hc/en-us/articles/360037003314) for Glossaries. |
| **Lineage** | Diagram on the data object page that helps trace the origin of data in Tables and Columns and gives visibility into how data in a source moves and transforms over time. Input data for Lineage diagrams comes from QLI, queries in Compose, and can be added via the dedicated API. |
| **Metadata Extraction (MDE)** | Extraction of metadata from a data source into the Catalog with the goal of synchronizing them and providing Alation users with a representation of the data contained in this source. [Metadata extraction](https://alationhelp.zendesk.com/hc/en-us/articles/360038689193) can be run manually or on an automatic schedule. |
| **NoSQL Database** | NoSQL databases are software systems that provide storage and query capabilities to support specific data models and access patterns for building modern applications using flexible schemas. |
| **Object Template** | Catalog page [template](https://alationhelp.zendesk.com/hc/en-us/articles/360038214553) for Alation objects, such as data objects and queries. Can be modified by adding/removing custom fields. Cannot be created or removed by Alation users. |
| **Obfuscate Literals** | When enabled, hides all literals in queries. Only the query owner will be able to see the full query. |
| **Object Type** | Property of an Alation object that describes what kind of object it is. Each specific Alation object has a type. For example, your MySQL data source is an object of type "**data**", or your article "How to use JOINs in Alation Analytics Database" is an object of type "**article**". You can get a list of all object types we use by querying Alation Analytics: [Understanding Alation Analytics: Object Queries](https://alationhelp.zendesk.com/hc/en-us/articles/360016445494). |
| **Popularity** | An indicator of the intensity of use of a data object. Popularity is determined with an algorithm that accounts for both query ingestion (the number of data source accounts that used the object and the number of times the object was used) and Compose usage (the number of Alation users that used the object in Compose and the number of times the object was used.) |
| **Profiling/Sampling** | Creation of data samples for the cataloged data sources. When you perform data profiling in Alation, it scans the first 10,000 rows of a table. A data sample is a random sample of 100 rows out of the 10,000 scanned rows intended to provide you an insight to what kind of data is contained in the profiled table without the need to run a query. |
| **Steward** | See **Data Steward**. |
| **Query Collaboration** | A collaborative workflow in Compose that lets users develop queries together. You can co-create queries leveraging the expertise of all Authors and the tribal knowledge. A query Owner can start the collaboration workflow by adding Authors to their queries. All Authors can take turns editing the query statement. They can also invite more Authors to contribute. |
| **Query Execution** | The action and process of running a query using Alation. You may also come across the term "query run" which means the same. |
| **Query Log Ingestion (QLI)** | The processing of the database query logs in order or to calculate Popularity, Lineage, and Top Users, and to supply filter and JOIN information to the catalog. |
| **Query Result Table** | The results of the query execution. You can save the results of a query as the Result Table object in the Alation catalog. |
| **Server Admin** | A role in Alation. Has access to the full scope of functionality in the Catalog and Compose, including the application settings. |
| **Social Querying** | See **Query Collaboration**. |
| **Standard User** | A role in Alation. Has access to Compose and the main features of the Catalog, with the exception of ddmin settings and catalog customization tools (such as templates and custom fields). |
| **Top Users** | On Schema and Table catalog pages, Alation automatically lists **Top Users** through calculating expert status. Alation calculates expertise on a data object based on a number of factors: Number of the user’s queries that mention the data object, Number of times the user added comments, titles, or descriptions to the data object page, Nomination by other users. |
| **Tree Table** | An table-like element of the catalog interface with expandable rows revealing nested elements. Mostly used to represent complex data types, such as records, arrays, or structs. |
| **Virtual Data Source** | A type of source object in Alation that can be used to catalog a data storage that is either unsupported by Alation with a built-on or custom driver or for some reason cannot be accessed from Alation using the URI (could be for security reasons). There is no automated MDE, Profiling and Sampling, or QLI for Virtual Sources, and the metadata should be uploaded via the API. One example of the use of a Virtual Data Source is to catalog Avro schemas from data streaming platforms. Virtual Sources can be used for querying in Compose if the database supports SQL dialects and if you choose to add the Compose connection information. |