



ThoughtSpot Administration Guide

Release 6.3

November, 2020

© COPYRIGHT 2015, 2020 THOUGHTSPOT, INC. ALL RIGHTS RESERVED.

910 Hermosa Court, Sunnyvale, California 94085

This document may not, in whole or in part, be copied, photocopied, reproduced, translated, or reduced to any electronic medium or machine-readable form without prior consent in writing from ThoughtSpot, Inc.

All rights reserved. The ThoughtSpot products and related documentation are protected by U.S. and international copyright and intellectual property laws. ThoughtSpot and the ThoughtSpot logo are trademarks of ThoughtSpot, Inc. in the United States and certain other jurisdictions. ThoughtSpot, Inc. also uses numerous other registered and unregistered trademarks to identify its goods and services worldwide. All other marks used herein are the trademarks of their respective owners, and ThoughtSpot, Inc. claims no ownership in such marks.

Every effort was made to ensure the accuracy of this document. However, ThoughtSpot, Inc., makes no warranties with respect to this document and disclaims any implied warranties of merchantability and fitness for a particular purpose. ThoughtSpot, Inc. shall not be liable for any error or for incidental or consequential damages in connection with the furnishing, performance, or use of this document or examples herein. The information in this document is subject to change without notice.

Table of Contents

Introduction to administration	7
Sign-in credentials for administration	8
Understand the architecture	
Architectural components	11
Data caching.....	12
Authentication frameworks.....	14
Data and object security.....	15
Performance considerations.....	18
In-memory data compression.....	20
Admin Console	
Overview	23
Managing users.....	25
Managing groups	27
Local authentication	29
Authentication through SAML.....	31
Authentication through Active Directory	34
Configure SSL.....	40
Configure a reverse SSH tunnel.....	46
Set the relay host for SMTP (email)	49
Customize ThoughtSpot Help	52
Customize actions.....	56
Style Customization	59
Configure NAS for backup storage	61
Manage and create snapshots.....	65
System Cluster Pinboard	69
System Alerts Pinboard	73
User Adoption Pinboard	75
Available cluster updates	77
Installation and setup	
About installation and upgrades	78
Set your locale	81
Test connectivity between nodes	84

ThoughtSpot use agreement	85
Set the relay host for SMTP	89
Set up custom calendars	94
Configure internal authentication	99
Configure SSL	102
Configure SAML	109
Enable SSH through Active Directory	115
Integrate LDAP	
About LDAP integration	117
Configure authentication through Active Directory	118
Add the SSL certificate for LDAP	125
Test the LDAP configuration	126
Sync users and groups from LDAP	127
Configure NAS file system	132
Set up monitoring	137
Configure support services	139
Network ports	148
Configure load balancing and proxies	160
Customize ThoughtSpot Help	162
Customize look and feel	166
Load and manage data	
Introduction to data management	170
Configure casing	172
Load CSV files with the UI	173
How to view a data schema	178
Plan the schema	
About schema planning	184
Data types	187
Constraints	191
Sharding	196
Chasm traps	208
Build the schema	
Schema building overview	211
Connect with TQL and create a schema	213
Create a schema in SQL	215
Examples of schema creation	217

Upload a SQL script.....	222
Change the schema	
How to change a schema.....	228
Convert column data type	233
Load bulk data	
Import CSV files with tsload.....	238
Use a script.....	247
Use the tsload connector	252
Delete a data source	
Delete a data source (table)	258
Delete or change a table in TQL	262
Manage users and groups	
Onboarding Users	263
Understand groups and privileges	265
Create, edit, or delete a group	271
Create, edit, or delete a user.....	287
Allow users to sign up.....	297
Security	
Overview of security features	299
System security	
Tools and processes.....	300
About third-party software.....	303
Installing third-party software	306
Data security	
Data security.....	308
Share tables and columns.....	314
Share worksheets	318
Share Views	322
Share Pinboard.....	326
Share Answers	336
Revoke access (unshare)	346
Security for SpotIQ functions.....	348
Row level security (RLS)	
About row level security (RLS).....	349
How Rule-Based RLS works	351

Set Rule-Based RLS.....	356
ThoughtSpot Lifecycle	359
Encryption of data in transit.....	365
System administration	
Overview of System administration.....	367
Send logs when reporting problems	368
Set up recording for Replay Search	370
Upgrade a cluster.....	375
Backup and restore	
Understand the backup strategies	377
Understand backup/snapshot schedules	380
Work with snapshots	385
Work with backups	
Understand backup modes.....	393
Create a manual backup	396
Configure periodic backups.....	399
About restore operations	403
Improve search with modeling	
About data modeling.....	404
Change a table's data model	406
Edit the system-wide data model	408
Data model settings	
Overview of the settings.....	413
Set column name, description, and type	416
Set additive and aggregate values	420
Hide a column or define a synonym.....	424
Set columns to exclude from SpotIQ analyses	426
Manage suggestion indexing	428
Add a geographical data setting	434
Set number, date, currency formats	438
Change the Attribution Dimension setting	444
Add or manage experts	447
Link tables using relationships	
Link tables using relationships	449
Create a relationship	450

Delete a relationship.....	455
Use stickers.....	457
Worksheets: improve Search	
Create and use worksheets	461
Edit a worksheet.....	467
Create a formula in a worksheet	469
Create worksheet filters	472
How the worksheet join rule works.....	477
Change join rule or RLS for a worksheet.....	481
Create a join relationship	483
Modify joins between Worksheet Tables	488
Delete Worksheets or Tables	491
Work with views	
Understand views.....	494
Save a search as a view.....	497
Create a search from a view	499
View example scenarios.....	501
About materialized views	503
Materialize a view.....	506
Dematerialize a view	513
Refresh a view	515
Schedule view refreshes.....	517
Scriptability	
Overview	520
Migrate or restore objects.....	521
ThoughtSpot Scripting Language.....	535
SpotApps.....	556
Set up SearchIQ	
Enable SearchIQ.....	560
Optimize SearchIQ	562
Search IQ Optimization Steps	
Enabling Columns in SearchIQ	564
Add Experts for SearchIQ.....	570
Train SearchIQ.....	573
Set entity categories for SearchIQ	577

Change SearchIQ mappings	579
Manage scheduled jobs	
Job management (scheduled pinboards)	582
Scheduled pinboards management.....	585
Monitoring	
Introduction to monitoring	587
System Information and Usage Pinboard	589
Data board	603
Cluster Manager board.....	606
Alerts and Events board	609
System worksheets.....	611
System Pinboards	613
Falcon monitoring Pinboards.....	616
Performance Tracking Pinboard.....	621
Troubleshooting	
About troubleshooting	623
Get your configuration and logs	625
Upload logs to ThoughtSpot Support	630
Network connectivity issues	632
Check the timezone	633
Browser untrusted connection error.....	634
Characters not displaying correctly	635
Clear the browser cache	636
Cannot open a saved answer that contains a formula	639
Data loading too slowly	642
Search results contain too many blanks.....	643

Introduction to administration

Summary: This guide covers all topics of special interest to application administrators.

This guide provides information for application administrators, or users with [administrative access privileges](#) [See page 8].

Before addressing the major components of this guide, we recommend that you familiarize yourself with the general top-level [architecture](#) [See page 11] of the ThoughtSpot service.

Administrators are responsible for many facets of the ThoughtSpot service. They are most frequently in charge of these common processes:

- [Installation and setup of ThoughtSpot](#) [See page 78]
- [Loading and managing data](#) [See page 170]
- [Managing users and groups](#) [See page 265]
- [Security](#) [See page 299]
- [System administration](#) [See page 367]
- [Backup and Restore](#) [See page 377]

Additionally, administrators are often involved in the following workflows:

- [Data modeling](#) [See page 404]
- [Using worksheets](#) [See page 461] to simplify search
- [Using views](#) [See page 494] for ‘stacked’ search; note that starting with Release 5.2, you can accomplish some aspects of search stacking by using the `IN` keyword (See <https://docs.thoughtspot.com/6.0/complex-search/in-keyword-searches.html>).
- [Enabling SearchIQ](#) [See page 494] Beta, ThoughtSpot’s natural language search.
- [Managing scheduled jobs](#) [See page 582]
- [Monitoring system health](#) [See page 587]
- [Troubleshooting](#) [See page 623]

Sign-in credentials for administration

Summary: You must have administrative access to perform various administrative tasks.

You can access ThoughtSpot through SSH at the command prompt, and from a Web browser.

Administrative access

Each ThoughtSpot cluster has three default users. Contact your ThoughtSpot support team to get the passwords.

Type	Username	Description
Shell user	admin	<p>For work that requires <code>sudo</code> or <code>root</code> privileges</p> <p>Not for application login</p> <p>Logs for this user are in <code>/usr/local/scaligent/logs</code> directory</p>
Shell user	thoughtspot	<p>For command-line work that does not <code>sudo</code> or <code>root</code> privileges</p> <p>Can use <code>tsload</code>, <code>tql</code>, and check the cluster status</p> <p>Not for application login</p> <p>Logs for this user are in <code>/tmp</code> directory</p>
Application user	tsadmin	Access through a Web browser

Both the `admin` [See page 8] and `thoughtspot` [See page 8] user can SSH into the cluster. After authenticating, either user can use and all of the following utilities:

- `tscli` [See page 0]; `thoughtspot` [See page 8] user cannot use commands that require `sudo` or `root` privileges
- `tsload` [See page 238]
- `tql` [See page 0]

SSH to the appliance

To perform basic administration such as checking network connectivity, starting and stopping services, and setting up email, log in remotely as the Linux administrator user “admin”. To log in with SSH from any machine, you can use the command shell or a utility like Putty.

In the following procedure, replace `<hostname_or_IP>` with the hostname or IP address of a node in ThoughtSpot. The default SSH port (22) will be used.

1. Log in to a client machine and open a command prompt.
2. Issue the SSH command, specifying the IP address or hostname of the ThoughtSpot instance:

```
ssh admin@<hostname_or_IP>
```

3. Enter the password for the admin user.

ThoughtSpot provides a default password for both the `admin` and `thoughtspot` shell users. ThoughtSpot recommends that you change the passwords for both these users on first login. Follow your company’s policies on password security.

Sign in to the ThoughtSpot application

To set up and explore your data, access the ThoughtSpot application from a standard Web browser, using a username and password.

Before accessing ThoughtSpot, you need the following:

- The Web address (IP address or server name) for ThoughtSpot
- A network connection
- A Web browser
- A username and password for ThoughtSpot

ThoughtSpot supports the following Web browsers:

Firefox

68.x, 69.x, and later

Chrome

76.x, 77.x, and later

Internet Explorer

11.x, and later

Edge

81.0.416.53, and later

Safari

13.x, and later

☒ Tip: We support, but do not recommend, the use of the Internet Explorer.

Depending on your environment, you can experience performance or UI issues.

To sign in to ThoughtSpot from a browser, follow these steps:

1. Open the browser and type in the Web address for ThoughtSpot:

`http://<hostname_or_IP>`

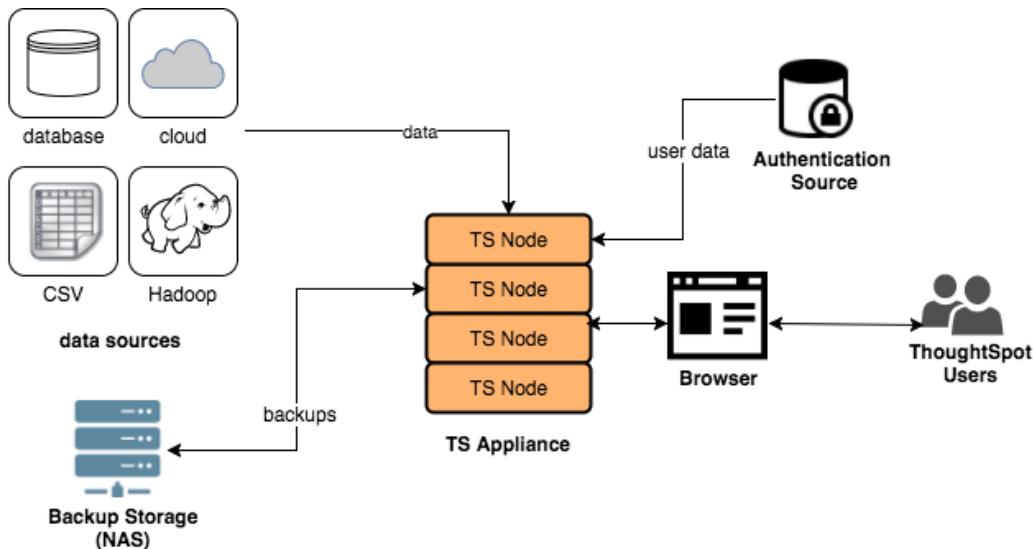
2. Enter your username and password, and click **Sign in**.

Architecture components

Summary: To implement ThoughtSpot it is important to understand where it sits within your overall analytics architecture and how it provides data to end users.

ThoughtSpot consists of a cluster of one or more nodes, acting together to provide analytic answers to business questions. As such, there are only a few integration points with ThoughtSpot on your network.

The major components in a ThoughtSpot cluster are:



ThoughtSpot can handle a wide variety of different data sources. ThoughtSpot does all analysis against data in memory to help achieve fast results across millions and billions of records of data. ThoughtSpot caches the data in order to process it.

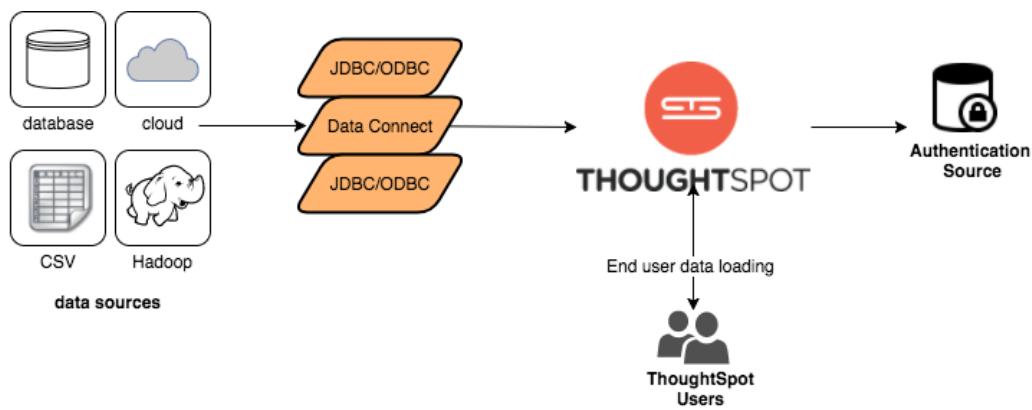
The ThoughtSpot appliance can be a physical appliance that ThoughtSpot ships, one or more AWS instances that are clustered together, or one or more VMware instances that are clustered together. From an external interface, regardless of the appliance type, the appliance appears to be a single instance.

For authentication (logging in), some source of user information is required. These define the login requirements and access control groups. Users must access the data from a supported browser to view saved content, or perform search-based analytics. Finally, it is recommended that you have some sort of networked attached storage for storing backups in case of hardware failure.

Data Caching

Summary: ThoughtSpot does all analysis against data in memory to help achieve fast results across millions and billions of records of data.

ThoughtSpot caches data as relational tables in memory. The tables can be sourced from different data sources and joined together. ThoughtSpot has several approaches for getting data into the cluster.



JDBC and ODBC Drivers

ThoughtSpot provides a JDBC and ODBC driver that can be used to write data to ThoughtSpot. This is useful for customers who already have an existing ETL process or tool, and want to extend it to populate the ThoughtSpot cache.

JDBC and ODBC drivers are appropriate under the following circumstances:

- have an ETL load, such as Informatica, SSIS, and so on
- have available resources to create and manage ETL
- have smaller daily loads

tsload

You can use the `tsload` command line tool to bulk load delimited data with very high throughput.

Finally, individual users can upload smaller (< 50MB) spreadsheets or delimited files.

We recommend the tsload approach in the following cases:

- initial data load
- JDBC or ODBC drivers are not available
- there are large recurring daily loads
- for higher throughput; this can add I/O costs

Choosing a Data Caching Strategy

The approach you choose depends on your environment and data needs. There are, of course, tradeoffs between different data caching options.

Many implementations use a variety of approaches. For example, a solution with a large amount of initial data and smaller daily increments might use tsload to load the initial data, and then use the JDBC driver with an ETL tool for incremental loads.

Authentication

Summary: ThoughtSpot provides LDAP/AD, SAML, and ThoughtSpot login to authenticate users.

ThoughtSpot provides three ways to authenticate users: LDAP/AD, SAML, and ThoughtSpot login. If possible, ThoughtSpot recommends that you use LDAP/AD or SAML, since ThoughtSpot provides only basic authentication, with no restrictions on passwords, timeouts, failed logins, etc.

Use the following table to help you choose an authentication option.

SAML [See page 109]	LDAP/AD [See page 117]	ThoughtSpot [See page 287]
<ul style="list-style-type: none"> • Use SAML for single sign-on authentication. • Can redirect from ThoughtSpot to SAML logins. • Recommended for portal integration. • Option to sync users and groups if stored in LDAP/AD. 	<ul style="list-style-type: none"> • Configuration. • Users authenticate against LDAP or AD. • Option to sync users and groups with ThoughtSpot to manage group membership. 	<ul style="list-style-type: none"> • User created and managed in ThoughtSpot. • Password strength control [See page 290] • No other enterprise password control (expiration, failed logins, etc.). • Only recommended when SAML and LDAP are not options.

All users and groups must be known to ThoughtSpot. If you are using LDAP/AD or SAML and don't create users in ThoughtSpot, a user is created when the user first logs in. However, this user is assigned to the `All` group and can only see content available for all users.

Groups are the primary way that security is managed. Groups are not automatically created. You can create [groups \[See page 271\]](#) and [users \[See page 287\]](#) manually, or you must automate the assignment from a source system. ThoughtSpot has an assignment script that works with most LDAP / AD stores. It also has public APIs that you can use to sync users and groups between source systems and your ThoughtSpot appliance.

Data and object security

Summary: ThoughtSpot provides many features for protecting data.

Object Security

Object security controls what content users see within ThoughtSpot. Objects are tables, columns in tables, worksheets, pinboards, and saved answers.

Users gain access to objects when an object owner shares access with them. Owners can share with individual users or with entire groups, giving access to everyone within that group. Objects may be shared with edit or view-only options. A user can automatically share objects with anyone else in the groups to which they belong. This has implications on setting up privileges, and on applying row-level security.

Permissive Security Mode

The default Permissive Security mode of ThoughtSpot means that when someone shares an object with you, you can see all the data it uses, regardless of explicit permissions to the parent object data. You can see a shared pinboard without having access to its underlying worksheet or table.

Advanced Security Mode

ThoughtSpot's Advanced Security mode is opposite of the default permissive mode. Unless the user has explicit permissions to the entire stack of parent objects, they cannot see the data in the child object. For example, in a shared pinboard, you can see data only if you have explicit permissions to the relevant columns of the parent worksheet. Similarly, you can only see the data in a worksheet to which you have access if you have explicit permissions to its parent table object.

Work with your ThoughtSpot support team to enable the Advanced Security Mode on the relevant clusters.

Row level security (RLS)

Row level security controls what data a user can see in each shared piece of content. Even if a user has access to a worksheet, they can only see rows from the tables they have permission to see.

RLS applies at the table level, so it automatically extends to all worksheets, saved answers, and pinboards based on that table, every time. Also, in queries where there are tables with table filters, all joins are always enforced to avoid accidentally allowing users access to data they shouldn't see.

RLS requires three things:

- A table filter with a column (possibly in a joined table) that can be used to determine who can see a row, such as account id or tenant id.
- A group that can be associated with the row of data by name. For example, if the column is `account_id` and has values of `1`, `2`, `3`, users can be assigned to groups `group_1`, `group_2`, `group_3` and then only see their data.
- Users must be assigned to the group. If they are not assigned to a group that has access, they do not see any data.

Administrative users can always see all rows of data because RLS does not apply to them.

RLS supports a hierarchy of groups, which makes it possible to grant access to some users across multiple groups.

Keep in mind that users within a group can share with one another. If you put everyone in your organization into the same group for RLS, they can share with anyone in the company.

Column level security (CLS)

Column level security lets users see certain columns in a table, but not other columns. This can be accomplished by sharing a limited set of columns in a table with specific users or groups.

Because someone can share with anyone in the same group, they can potentially share restricted columns. For example, if a *Human Resources* repository has a column with salary information, and it appears in a worksheet, any *Human Resources* group member could create an answer with visible salary information and mistakenly share with someone outside of *Human Resources*. That ‘outside’ person now has access to the salary information. In such cases, we recommend that you work with your ThoughtSpot support team to enable the Advanced Security Mode on the relevant clusters.

System privileges

System privileges refer to what a user can do in ThoughtSpot. For example, can they upload or download data or share with all users. These privileges are defined on a group level and inherit downwards. So, if Group A had child groups Group B and Group C, then any privilege given to Group A is also available to Group B and Group C. What this often means is that separate sets of groups are required to manage privileges.

Performance considerations

Summary: Make sure you understand the performance considerations in your installation.

ThoughtSpot configuration and licensing varies by memory availability. Other considerations also impact the performance of your solution. Because some solutions perform better than others, think about the following issues before implementation.

Nodes that have 250GB of memory capacity perform optimally with less than 250GB of data, and less than 250 million rows of data in each node in a ThoughtSpot cluster. Smaller nodes, like nodes with 200GB memory capacity, serve proportionally smaller loads.

To optimize performance for more complex schemas, we recommend fewer rows of data for each node. To reduce the total amount of data and rows of data, we recommend the following approaches:

- Limit the data range to the relevant years or months.
- Combine long and narrow tables into wider tables whenever possible.

Data Boundaries

Total rows in a result of a join can have an impact on performance. In general, we recommend that you have fewer than 10 billion rows in a many-to-many join. Also, consider these boundaries:

Contact ThoughtSpot support for guidance on boundaries for the following:

- Maximum number of rows that can be downloaded
- Size in CSV format
- Total number of rows across all tables
- Many-to-Many (Generic join cardinality)
- Load frequency

Worksheet Boundaries

Worksheets must have less than 1000 columns. For aggregated worksheets, follow these guidelines:

- Number of columns should be less than 50

- Number of rows should be less than 10 millions

You can use an ETL process to circumvent these limitations. Speak with ThoughtSpot support to learn more.

Aggregated Worksheets and Joins

To join an aggregated worksheet with a base table, you must configure your installation to allow this behavior.

- The aggregated worksheet cannot use more than 5 component tables.
- The number of rows in the final aggregated worksheet cannot be greater than 1000.

Chasm Trap Worksheets

For chasm trap scenarios where two or more fact tables join through a shared dimension, we recommend the following boundaries:

Description	Boundary
Maximum number of fact tables in a worksheet	5
Maximum number of shared dimensions	2
Maximum number of rows in a <i>non</i> co-sharded shared dimension table of chasm trap	1B
Maximum number of rows in a co-sharded shared dimension table of chasm trap	1B

Row-level Security Boundaries

Maximum number of unique RLS rules with search data suggestions should not exceed 15K.

Scheduled Pinboards

For ideal performance of scheduled pinboards, do not exceed 50 scheduled pinboard jobs.

In-memory data compression

Summary: In-memory data compression lowers RAM requirements and reduces the number of VMs you need for in-memory data, reducing deployment costs for appliance and cloud platforms.

The cost to deploy ThoughtSpot on an appliance or cloud platform depends on your data capacity.

ThoughtSpot release 6.2 includes improvements to in-memory data compression. These improvements lower your RAM requirements and reduce the number of VMs you need for in-memory data, when deploying on a cloud platform.

In release 6.2, ThoughtSpot added two new compression algorithms to the Dictionary compression that ThoughtSpot already supports. ThoughtSpot now supports **LZ4**, for `INT`, `BIGINT`, `DOUBLE`, and `FLOAT` data types, and **RLE**, for strings.

These improvements to in-memory data compression may result in up to 20-50% compression of your in-memory data, reducing your cloud and appliance costs.

In-memory data compression is turned on by default for new clusters on release 6.2. If you upgrade from an older release to 6.2, this feature is **not** on by default. After you upgrade to 6.2, ThoughtSpot compresses any new tables you create, but does not compress your existing tables. To enable compression of your existing in-memory data after you upgrade to 6.2, or to find out if it could lower your data costs, [contact ThoughtSpot Support](#) [See page 0].

Note: ThoughtSpot compresses **only** your fact tables, not your dimension tables.

View your in-memory data compression information

You can view compression ratios for your overall system, and for each table individually.

You can see the compression ratio from the **Relational Data Cache** panel under **Data > Usage > Data**. Here, the compression ratio is 1:1.34.

Relational Data Cache

LAST DATA UPDATE Aug 17, FY 2021 04:17:43 PM PDT

2572 TABLES LOADED

0 of 2572 TABLES BEING UPDATED

0 NEW TABLES BEING LOADED

1.1B ROWS

1:1.34 COMPRESSION RATIO

To view compression ratios for each table individually, navigate to the **Table Information** table. Click on the **Data** tab from the top navigation bar. Select **Usage > Data**. Scroll down to the **Table Information** table.

You can see the compression ratio from the **Compression Ratio** column, when you scroll all the way to the right on the table. Here, the compression ratio for the first table is 1:100.09.

The screenshot shows the ThoughtSpot Admin interface with the 'Usage' tab selected for the 'Embrace' dataset. The 'Compression Ratio' column is highlighted with a blue border.

Total Row Count	Replicated	Used CSV Capacity (MB)	Row Count Skew	Search Status	Last Index Build Duration	Total Shards	Compression Ratio
03:38:22 P... 3081687	N	355.61	1283	UPDATING INDEX	0h2m51s	48	1:100.09
03:38:07 P... 2784608	N	812.62	841	READY	0h14m29s	48	1:47.24
03:36:32 P... 15637235	Y	13958.41	0	READY	0h1m47s	1	1:1.79
03:35:29 P... 17620	N	54.6	4106	NOT READY		48	1:1.92
03:35:26 P... 86580	N	12.14	28860	NOT READY		48	1:2.50
03:31:36 P... 8775609	Y	853.65	0	NOT READY		1	1:1
02:57:58 P... 172913	N	22.92	539	READY	0h1m47s	48	1:2.91
02:57:43 P... 72800	N	16.66	207	READY	0h1m47s	48	1:1.40
02:26:27 P... 51069	Y	51.14	0	NOT READY		1	1:2.42
02:26:17 P... 49	Y	0.03	0	NOT READY		1	1:1

Admin Console

Summary: Learn how the Admin Console can help you manage users, groups, security, authentication, and much more.

The Admin Console provides you with an intuitive, user-friendly interface to accomplish most of the necessary tasks for administering ThoughtSpot.

Navigate to the Admin Console by selecting **Admin** from the top navigation bar.



You can complete the following tasks in the Admin Console:

Manage users and groups

- [Manage ThoughtSpot users \[See page 25\]](#)
Create, edit, or delete users.
- [Manage ThoughtSpot groups \[See page 27\]](#)
Create, edit, or delete groups.

Configure authentication

- [Manage local authentication \[See page 29\]](#)
Enable or disable ThoughtSpot's local authentication.
- [Configure SAML authentication \[See page 31\]](#)
Integrate with SAML for single sign-on (SSO) authentication.
- [Configure LDAP authentication through Active Directory \[See page 34\]](#)
Integrate with LDAP to authenticate users against a secure LDAP server.

Manage settings: SSL, support, and application and menu customization

- [Configure SSL \[See page 40\]](#)
Use SSL for authentication and data security when you send data to and from ThoughtSpot.
- [Configure a reverse SSH tunnel \[See page 46\]](#)
Set up a reverse tunnel to allow ThoughtSpot Support to get access to your ThoughtSpot

instance, to perform support-related activities.

- [Set the relay host for SMTP \(email\) \[See page 49\]](#)

Set up a relay host for SMTP traffic, to route the alert and notification emails coming from ThoughtSpot through an SMTP email server.

- [Customize ThoughtSpot help \[See page 52\]](#)

Customize ThoughtSpot Help to be specific to your data, examples, and documentation.

- [Customize Answer actions menu \[See page 56\]](#)

Add custom actions to the more options menu  for Answers.

- [Customize application style \[See page 59\]](#)

Change the overall style of your ThoughtSpot interface: logo, colors, fonts, and footer text.

Manage backup and storage options

- [Configure NAS for backup storage \[See page 61\]](#)

Use network attached storage to support backup/restore and data loading.

- [Manage and create snapshots \[See page 65\]](#)

Create and manage snapshots, which you can use to restore the cluster to a specific point in time.

Monitor system

- [View System Cluster Pinboard \[See page 69\]](#)

View and explore cluster activity on the system cluster Pinboard.

- [View System Alerts Pinboard \[See page 73\]](#)

View and explore system alerts, configuration events, and notification events on the system alerts Pinboard.

- [View User Adoption Pinboard \[See page 75\]](#)

View and explore visualizations about how your ThoughtSpot users are interacting with ThoughtSpot, and how your user adoption is changing over time on the User Adoption Pinboard.

Upgrade your cluster

- [Monitor available cluster updates \[See page 77\]](#)

View available cluster updates.

Managing Users

Summary: Manage users in ThoughtSpot.

Before people can log in and use ThoughtSpot, you need to create a username, a password, and a membership in one or more groups for them. Use the Admin Console to easily create and manage your users.

Refer to [Understand groups and privileges \[See page 265\]](#) to learn about the privileges you can assign to a user.

To manage users, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Users** from the side navigation bar that appears.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A tsadmin	tsadmin	8 years ago

You can use the Admin Console to accomplish the following tasks:

- **Create a user** by clicking on the **+ add user** button
- **Manage an existing user** by clicking on the user you would like to edit: change the username, display name, sharing visibility, password, or email, or resend or test the welcome email
- **Delete a user** by hovering over a username, clicking the checkbox that appears, and selecting the **Delete** button
- **Add a user or users to groups** by hovering over a username, clicking the checkbox that

appears, and selecting the **Add users to groups** button

Refer to [Create, edit, or delete a user \[See page 287\]](#) for more detail on how to create and manage users.

Managing Groups

Summary: Manage user groups in ThoughtSpot.

Before people can log in and use ThoughtSpot, you need to create a username, a password, and a membership in one or more groups for them. Use the Admin Console to easily create and manage your user groups.

Refer to [Understand groups and privileges \[See page 265\]](#) to learn about the privileges you can assign to a user or group.

To manage groups, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Groups** from the side navigation bar that appears.

Display Name	Name ↑	Created	Default pinboards
Administration Group	Administrator	8 years ago	0
Group	Group	21 hours ago	0

You can use the Admin Console to accomplish the following tasks:

- **Create a group** by clicking on the **+ add group** button
- **Manage an existing group** by clicking on the group you would like to edit: change the group or display name, update the sharing visibility, add a description, edit the group's privileges, change the group's default Pinboard, add an existing group to the group, add a user to the group, or resend or test the welcome email
- **Delete a group** by hovering over a group name, clicking the checkbox that appears, and selecting the **Delete** button

Note: ThoughtSpot adds external users, or users that authenticate through SAML or Active Directory, to the **all** group by default. This group has no privileges. You must manually assign users to ThoughtSpot groups to give them privileges, such as **can upload user data**, or **can manage data**.

Related information

Refer to [Create, edit, or delete a group \[See page 271\]](#) for more detail on how to create and manage groups.

Local authentication

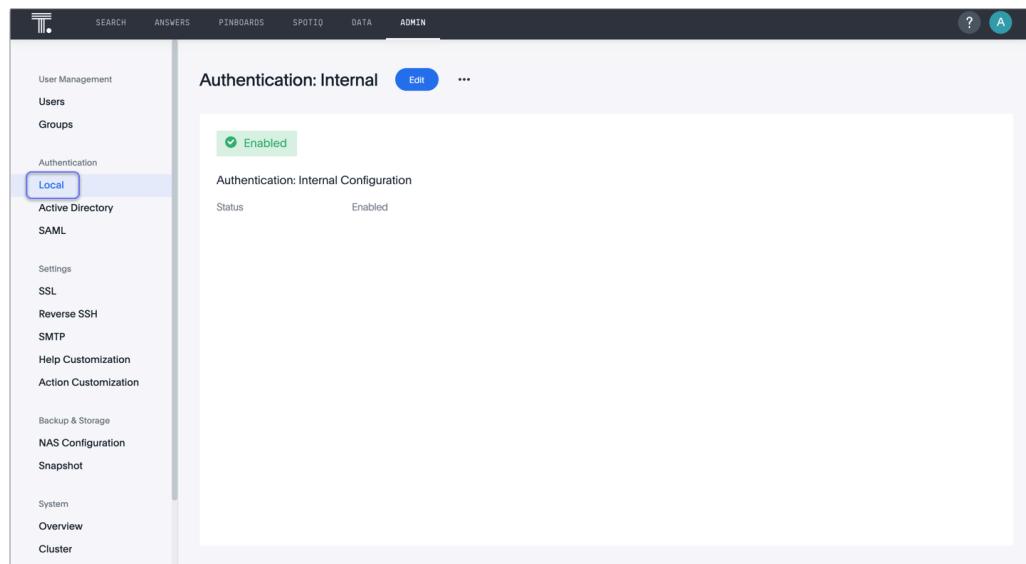
Summary: ThoughtSpot has robust local authentication.

Local authentication is enabled by default in your ThoughtSpot environment. This means that any local user you [create](#) [See page 25] can log into ThoughtSpot with the specific ThoughtSpot credentials you set up for them when creating the user.

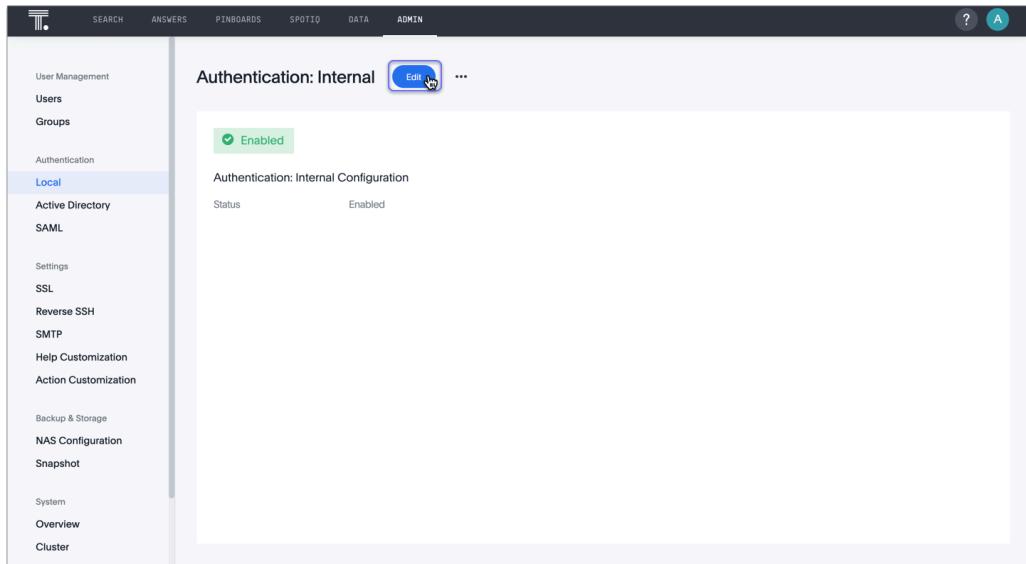
If you disable local authentication, no local user can log into ThoughtSpot.

You can enable or disable local authentication from the Admin Console.

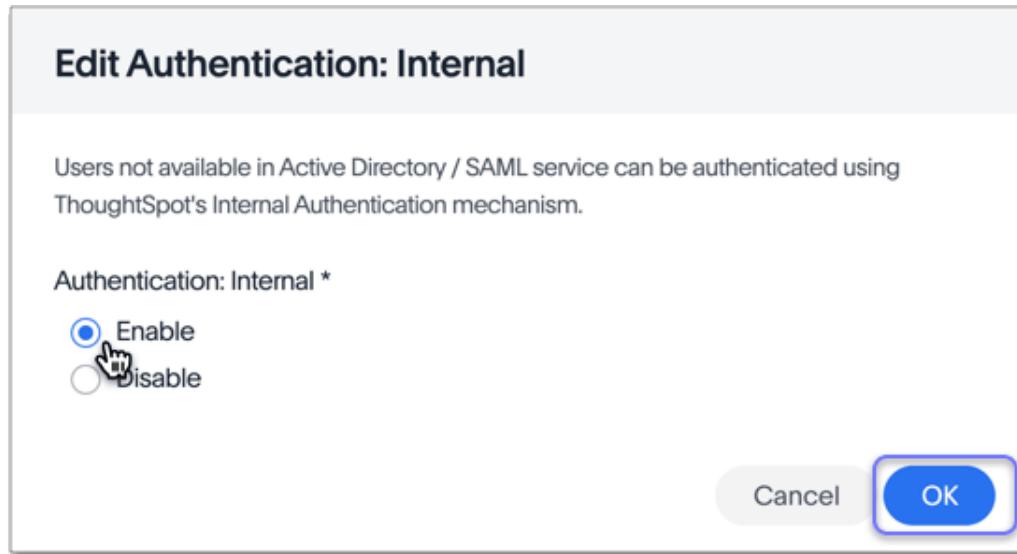
Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Local** from the side navigation bar that appears.



Click the **Edit** button in the top middle of the screen.



Select **Enable** or **Disable**, and click **OK**.



Note that you cannot disable local authentication until you configure [Active Directory \[See page 34\]](#) or [SAML authentication \[See page 31\]](#).

Authentication through SAML integration

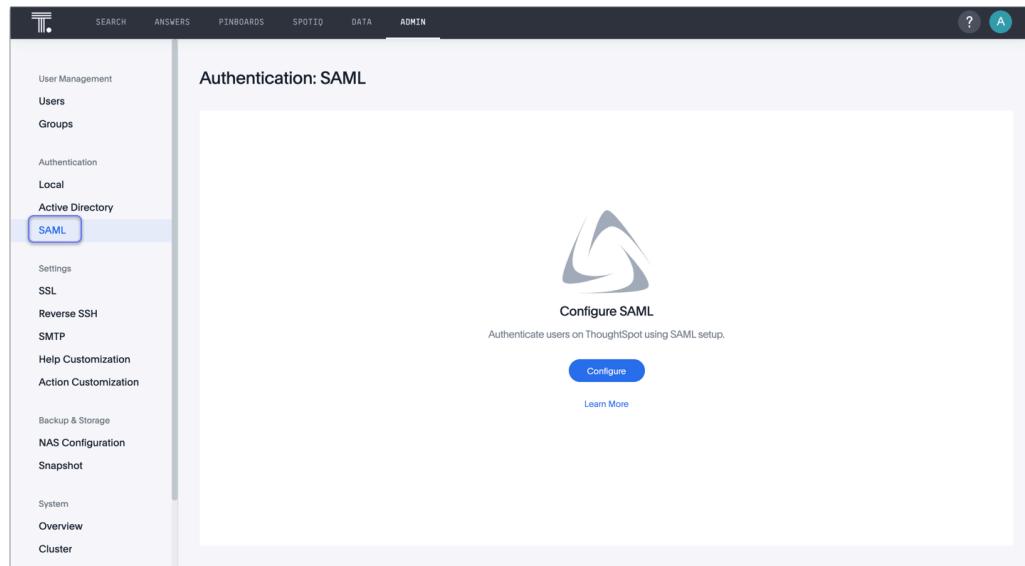
Summary: ThoughtSpot integrates with SAML for authentication.

You can use ThoughtSpot's integration with SAML for user authentication. By default, [local authentication](#) [See page 29] is enabled. You can also configure a SAML integration with an external Identity Provider (IdP), allowing your ThoughtSpot users to log in using one of the supported Identity Providers: Okta, Ping Identity, CA SiteMinder, or ADFS.

You can configure the SAML integration through the Admin Console.

Note: If you configure authentication through SAML, you cannot also configure authentication through Active Directory.

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SAML** from the side navigation bar that appears.



Click the **Configure** button in the middle of the screen.

Configure SAML

ThoughtSpot Service Address *
Supply fully qualified and resolvable domain name for ThoughtSpot service

Port *

Unique Service Name *
Unique service id will be used as IDP to identify client

Skew Time in Seconds *
Suggested skew time is 3600 seconds

Protocol *

IDP Metadata XML File *
[Browse File](#)

Automatically add SAML users to ThoughtSpot upon first authentication

Yes
 No

[Cancel](#) [OK](#)

Fill in the following parameters:

1. **ThoughtSpot Service Address:** A fully qualified and resolvable domain name for the ThoughtSpot service. For example, *thoughtspot.thoughtspot-customer.com*.
2. **Port:** Port of the server where your ThoughtSpot instance is running. For example, port `443`.
3. **Unique Service Name:** The unique key used by your Identity Provider to identify the client. For example, *urn:thoughtspot:callosum:saml*.

4. **Skew Time in Seconds:** The allowed skew time, after which the authentication response is rejected and sent back from the IDP. 86400 is a popular choice. The default is 3600.
5. **Protocol:** The authentication mechanism for ThoughtSpot. For example, `http` or `https`.
6. **IDP Metadata XML File:** The absolute path to your Identity Provider's metadata file. This file is provided by your IDP. You need this file so that the configuration persists over upgrades. It is a best practice to set it up on persistent/HA storage (NAS volumes) or in the same absolute path on all nodes in the cluster. For example, `idp-meta.xml`.
7. **Automatically add SAML users to ThoughtSpot upon first authentication:** Choose whether or not to add SAML users to ThoughtSpot when they first authenticate. If you choose 'yes', then new users will be automatically created in ThoughtSpot upon first successful SSO login. If you choose 'no', then SAML users will not be added in ThoughtSpot upon first successful SSO login. Instead, you must [add users manually](#) [See page 25].

After you fill in all parameters, click **OK**.

❶ Note: ThoughtSpot adds external users, or users that authenticate through SAML or Active Directory, to the **all** group by default. This group has no privileges. You must manually assign users to ThoughtSpot groups to give them privileges, such as **can upload user data**, or **can manage data**.

Authentication through Active Directory integration

Summary: ThoughtSpot enables you to set up integration with LDAP using Active Directory. After successful setup, you can authenticate users against a secure LDAP server.

You can use ThoughtSpot's integration with Active Directory for user authentication. By default, [local authentication \[See page 29\]](#) is enabled. You can also configure integration with LDAP through Active Directory, allowing you to authenticate users against an LDAP server. ThoughtSpot does not support any other LDAP authentication.

You can configure the Active Directory integration through the Admin Console.

Configuration prerequisites

Before you configure ThoughtSpot for Active Directory, collect the following information:

- URL [\[See page 34\]](#)
- Domain name [\[See page 35\]](#)
- Search base [\[See page 35\]](#)
- SSL [\[See page 35\]](#)
- Automatically add LDAP or AD users in ThoughtSpot? [\[See page 35\]](#)
- Also use ThoughtSpot internal authentication? [\[See page 0\]](#)

URL

Required to connect to Active Directory.

For example, `ldap://ad.yourdomain.local:389` or `ldap://ad.yourdomain.local:636`

Domain name

Default domain under which users who want to be authenticated against Active Directory reside. When a user logs in with a username, the default domain is added to the username before sending it to the LDAP server. If users reside in multiple sub-domains, you can still designate one of them as the default.

Authentication against multiple domains is not supported.

Users who don't belong to the default domain must explicitly qualify their username when they log in.

For example: `username@ad.yourdomain.local`

Search base

LDAP search base. The scope of searching user information, like *email* and *Display name*, within AD.

Automatically add LDAP or AD users in ThoughtSpot? (yes/no)

If you choose 'yes', new users are automatically created within ThoughtSpot when successfully authenticated against AD. ThoughtSpot doesn't cache passwords for AD-authenticated users.

If you choose 'no', users have to be manually created with a dummy password as a placeholder in ThoughtSpot before they can log in. The username you specify when creating the LDAP-authenticated user manually in ThoughtSpot has to be domain qualified, for example:

`username1@ad.yourdomain.local`.

In order to log in to ThoughtSpot, the user has to exist in ThoughtSpot independent of whether that user is authenticated against AD or against ThoughtSpot's internal authentication.

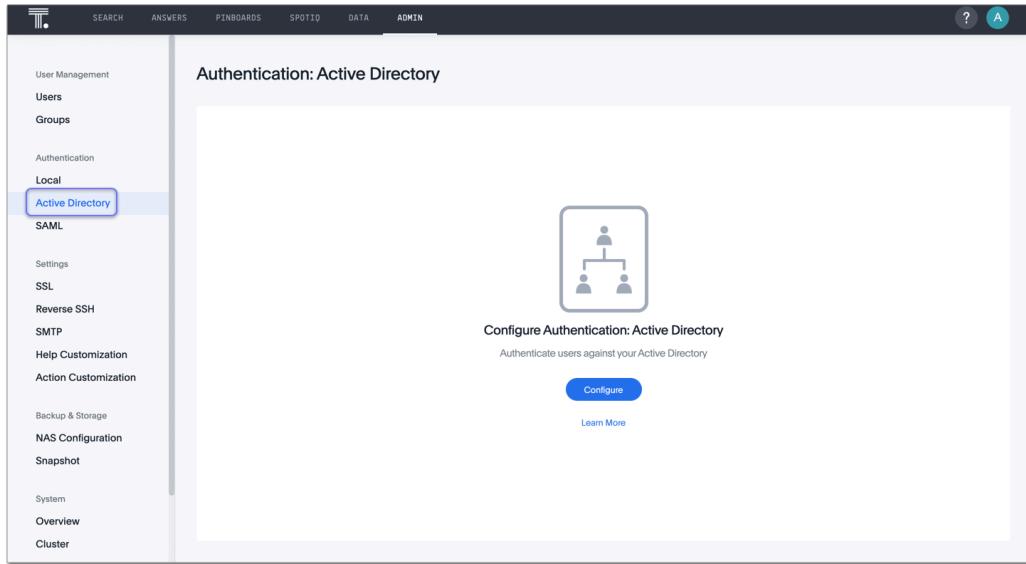
SSL

If you want to use SSL, you must obtain the SSL certificate from an issuing authority. When you select **yes**, ThoughtSpot prompts you to enter the certificate and certificate alias.

If AD servers are behind a load balancer, you must procure the SSL certificate to identify ThoughtSpot to the load balancer. The communication after the load balancer is non-secure. ThoughtSpot does not support a scenario where multiple AD servers provide their own SSL certificates.

Configure Active Directory

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Active Directory** from the side navigation bar that appears.



Click the **Configure** button in the middle of the screen, and add your information based on the information you collected in the [prerequisites \[See page 34\]](#).

Note: If you configure authentication through Active Directory, you cannot also configure authentication through SAML.

Configure Authentication: Active Directory

URL *

Domain Name Preferred *

Search Base *

Automatically add Active directory users to ThoughtSpot upon first successful authentication

Yes
 No

SSL Required

Yes
 No

Cancel OK

If you select **yes** for **SSL required**, ThoughtSpot prompts you to enter the SSL [certificate](#) [See page 35] and certificate alias. The certificate alias is a unique name you make up for your SSL certificate, in string format. If you replace the SSL certificate, you need a new certificate alias.

Configure Authentication: Active Directory

URL *

Domain Name Preferred *

Search Base *

Automatically add Active directory users to ThoughtSpot upon first successful authentication

Yes
 No

SSL Required

Yes
 No

SSL Certificate *

Could not pull SSL Certificate from AD server, upload certificate.

[Browse File](#)

Certificate Alias *

[Cancel](#) [OK](#)

After you add all your information, click **OK**.

Note: ThoughtSpot adds external users, or users that authenticate through SAML or Active Directory, to the **all** group by default. This group has no privileges. You must manually assign users to ThoughtSpot groups to give them privileges, such as **can upload user data**, or **can**

manage data.

Next steps

Next, sync users and groups from LDAP through Active Directory [See page 127].

Configure SSL

Summary: Secure socket layers (SSL) provide authentication and data security when sending data to and from ThoughtSpot.

You can use your own SSL certificate to secure ThoughtSpot HTTP(S) traffic.

About SSL

To enable SSL for the ThoughtSpot web service, you must provide your organization's SSL certificate for the ThoughtSpot service URL. If you do not have this certificate, you have the following options:

- Check with your IT department to see if they have an SSL certificate you can use.
- Obtain the certificate from an issuing authority.
- Use the default SSL certificate on the ThoughtSpot nodes.
- Disable SSL from the Admin Console: Under **SSL**, select **Configure > Disable or Edit > Disable**.

ThoughtSpot supports a wide variety of SSL types.

Required ports

To use the ThoughtSpot web service securely, ensure that TCP port 443 is open to accommodate incoming connections to ThoughtSpot nodes and clusters.

Prerequisites

To add SSL and enable HTTPS in ThoughtSpot, you must generate the [Certificate Signing Request \(CSR\)](#) [See page 40] and obtain the [SSL certificate chain](#) [See page 42] and the [private key](#) [See page 42].

Certificate Signing Request

When you generate a CSR, you handle sensitive data. Therefore, ThoughtSpot recommends that its customers generate their own CSRs.

You can generate a CSR in several ways. Most often, you generate a CSR and a new private key [at the same time \[See page 41\]](#). If you already have a private key, [use it to generate a CSR \[See page 41\]](#).

Follow these steps to generate a CSR and a private key. You need a computer you can run Linux commands on, and a recent version of `openssl`.

1. `ssh` into one of your ThoughtSpot nodes.

```
ssh admin@<node_IP>
```

2. Run the command to generate a CSR and private key pair:

```
openssl req -new -newkey rsa:2048 -nodes -out csr.pem  
-keyout pk.key [-subj "/key1=value1/key2=value with space/"]
```

Note the following parameters:

- ThoughtSpot supports a 2048 or 4096 bit key.
- `subj` : a common subject. Logically equivalent to the `-dname` property of `keytool`. Alternatively, you can skip this flag, and `openssl` prompts you to enter this information interactively.
- Optionally, run `add-multivalue-rdn` to allow multiple values to be set for the same key.
- Run `man req` for more details.

If you already have a private key, you can use it to generate a CSR. Follow these steps to generate a CSR with an existing private key:

1. `ssh` into one of your ThoughtSpot nodes.

```
ssh admin@<node_IP>
```

2. Run the command to generate a CSR and private key pair:

```
openssl req -new -key <private_key_file> -nodes -out csr.pem[-subj "/key1=value1/key2=value with space/"]
```

Specify the existing private key file. Refer to the parameters listed above.

SSL certificate chain

The SSL certificate chain must be in PEM format, which is an `X.509v3` file that contains ASCII (Base64) armored data, packed between `BEGIN` and `END` directives. The certificate chain may contain a series of certificates, with the root certificate at the bottom and user-facing, while the ThoughtSpot-specific SSL certificate is at the top.

Private key

The private key must also be in compatible PEM format. It cannot be password-protected, or passphrase-protected.

Note: Do not use a passphrase when creating certificates with ThoughtSpot.

If you are prompted to specify a passphrase, first check if it exists by invoking the following command:

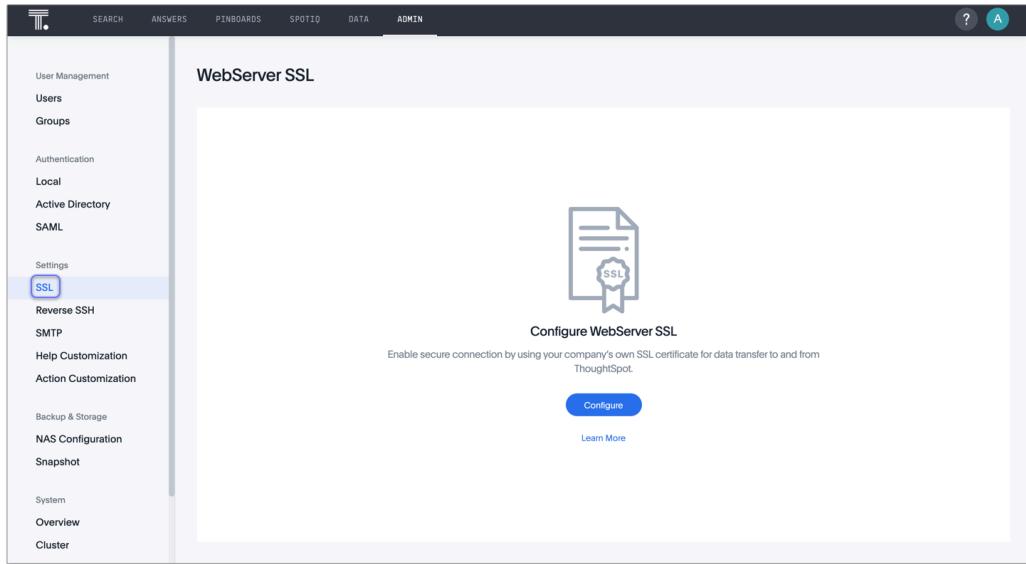
```
openssl rsa -check -in pk.key`
```

If the answer is ‘yes’, you must remove the passphrase first, and then proceed to use the private key with ThoughtSpot.

Configure SSL through the Admin Console

To configure SSL, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SSL** from the side navigation bar that appears.

Configure SSL



Click **Configure**.

Fill out the following parameters:

Configure WebServer SSL

SSL Status **1**
 Enable
 Disable

Choose RSA Algorithm * **2**
RSA

Choose Minimum TLS Version * **3**
TLS 1.2

Private Key * **4**
[Browse File](#)

SSL Certificate Bundle * **5**
[Browse File](#)

[Cancel](#) [OK](#)

1 Select **Enable**.

2 Choose **RSA**.

3 Choose **TLS 1.2** as a best practice. ThoughtSpot also supports **TLS 1.1** and **1.0**; set the minimum supported version to **1.1** or **1.0** to use these versions.

4 Attach your private key file.

5 Attach your SSL certificate bundle file.

Click **OK**.

Configuration string for load balancers

When enabling SSL support on a load balancer's server-side SSL client profile, make sure to add support for the following ciphers to ensure compatibility between the load balancer and ThoughtSpot.

The following ciphers are currently supported:

```
| TLSv1.2:  
|   ciphers:  
|     TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 - strong  
|     TLS_DHE_RSA_WITH_AES_256_CBC_SHA - strong  
|     TLS_DHE_RSA_WITH_AES_256_CBC_SHA256 - strong  
|     TLS_DHE_RSA_WITH_AES_256_GCM_SHA384 - strong  
|     TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384 - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 - strong  
|   compressors:  
|     NULL  
|_ least strength: strong
```

The cipher string would be as follows:

```
EECDH+AESGCM:EDH+AESGCM:AES256+EECDH: AES256+EDH
```

You can retrieve these from the ThoughtSpot web server (not against the load balancer) by running the following command on any ThoughtSpot node: `nmap --script ssl-enum-ciphers -p 443 <ThoughtSpot_node_IP_address>`

You must ensure that your load balancer supports these ciphers. If your load balancer cannot support these ciphers, [contact ThoughtSpot Support \[See page 0\]](#).

Test the SSL certificate

To test if the certificate is installed correctly, see [Sign in to the ThoughtSpot application \[See page 9\]](#).

Configure a reverse SSH tunnel

Summary: You can set up a reverse tunnel to allow ThoughtSpot Support to get access to your ThoughtSpot instance, to perform support-related activities.

Set up a reverse tunnel for support

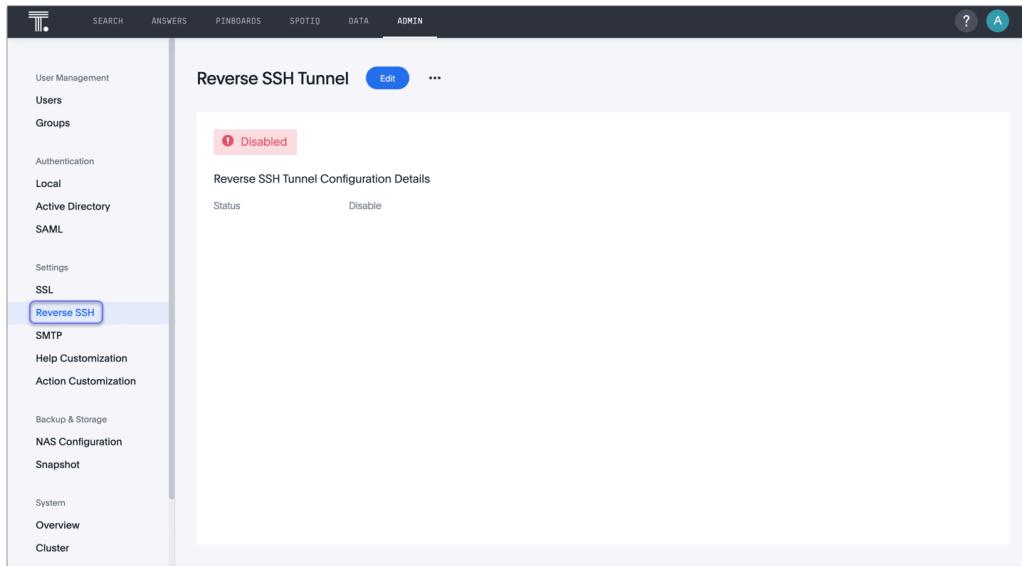
You can set up a reverse tunnel to allow ThoughtSpot Support to get access to your ThoughtSpot instance, to perform support-related activities. This setup is scalable, more secure, and a much simpler alternative to using a virtual meeting room.

Granting remote support access can streamline troubleshooting activities, because it enables your support agent to work directly on your computer from a secure setting. The remote tunnel enables SSH and HTTP access to your ThoughtSpot instance by ThoughtSpot Support. You can grant and revoke this access easily, so you can enable it for a troubleshooting session, and then disable it again. Before doing this procedure, make sure your company's security policies allow reverse tunneling.

Note: Before you set up a reverse tunnel, open port 22 in your firewall outgoing rules to add `tunnelrelay.thoughtspot.com` to your list of allowed domains.

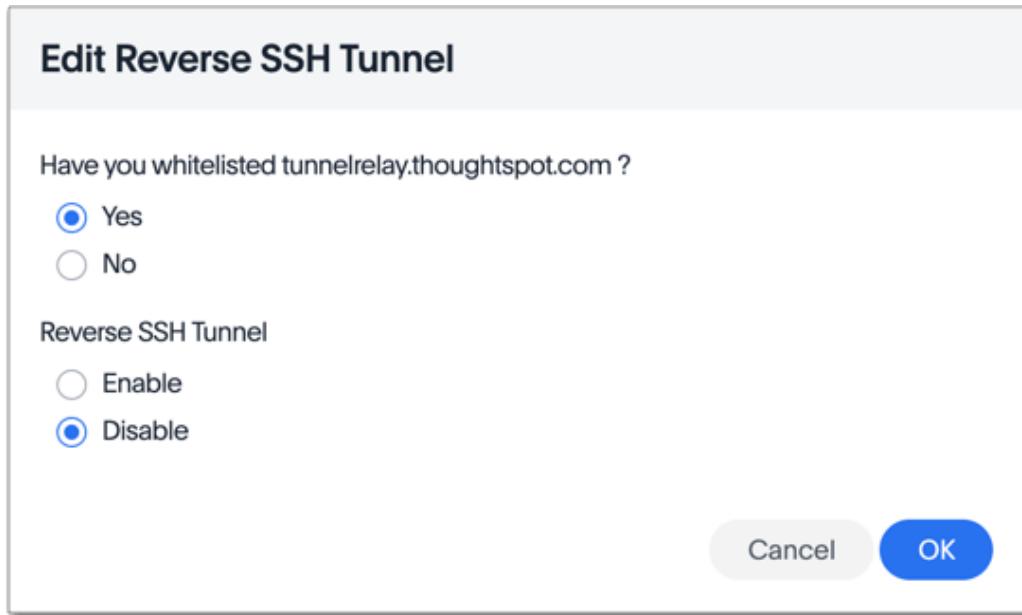
Configure reverse SSH tunnel

To configure a reverse SSH tunnel, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Reverse SSH** from the side navigation bar that appears.



Select **Edit** to enable or disable the tunnel.

The **Edit Reverse SSH Tunnel** dialog box appears.

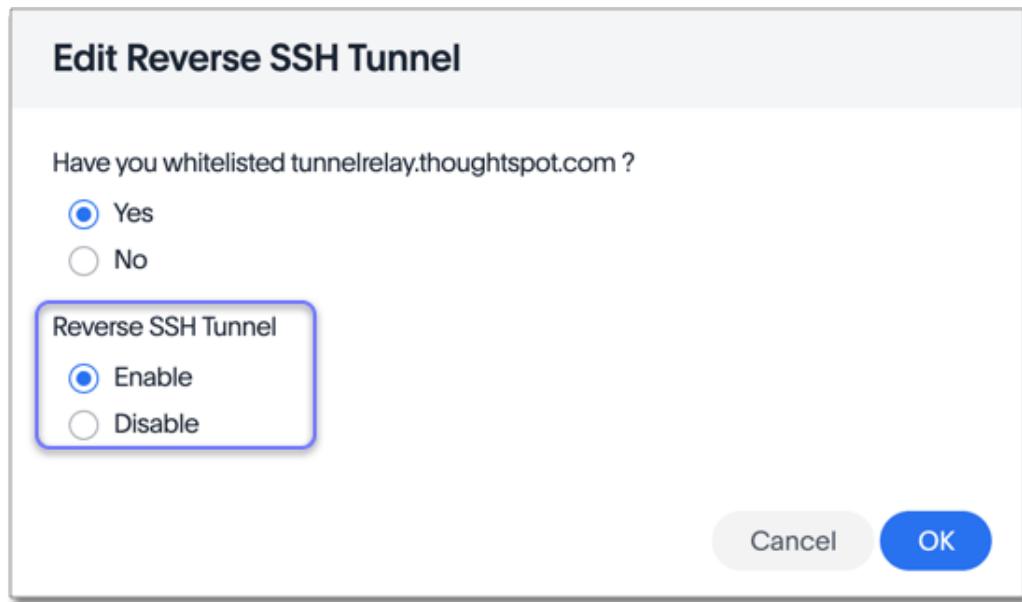


The system asks if you have added `tunnelrelay.thoughtspot.com` to your list of allowed domains.

If you select **no**, you cannot enable or disable the reverse SSH tunnel:



If you have added `tunnelrelay.thoughtspot.com` to your list of allowed domains, by opening port 22 , the system allows you to enable or disable the reverse SSH tunnel.



Make the necessary changes and select **OK**.

Set the relay host for SMTP (email)

Summary: ThoughtSpot uses emails to send critical notifications to ThoughtSpot Support. A relay host for SMTP traffic routes the alert and notification emails coming from ThoughtSpot through an SMTP email server.

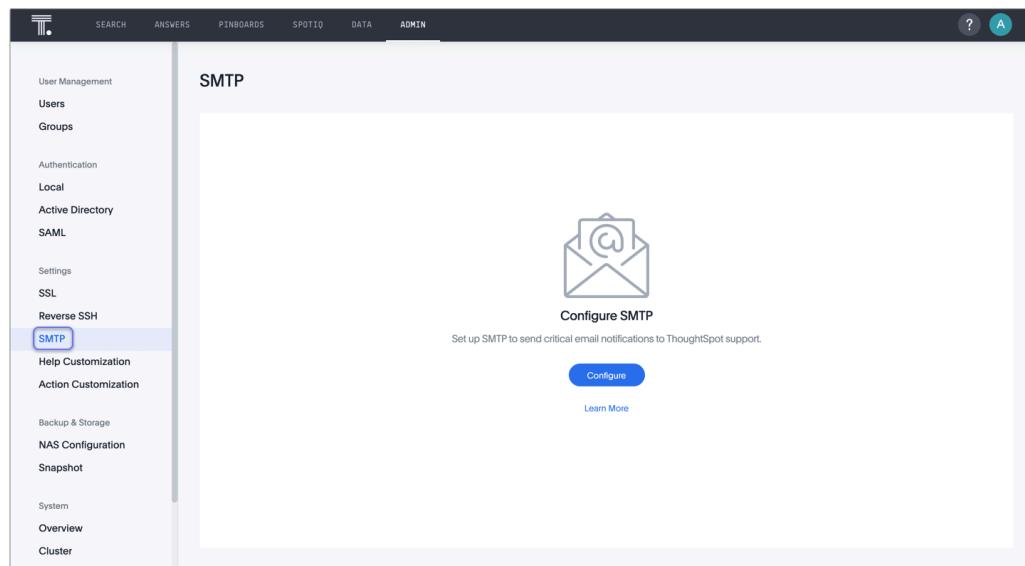
Configure SMTP through the Admin Console

You can set up the relay host for SMTP from the Admin Console.

Note: If you would like to use a custom port, rather than the default, port 25, you must configure SMTP using tscli, by running `tscli smtp set-relayhost <IP_address>:<custom_port>`.

Set up relay host

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SMTP** from the side navigation bar that appears.



Select **Configure**.

Specify the following parameters:

Configure SMTP

A relay host for SMTP traffic routes the alert and notification emails coming from ThoughtSpot through an SMTP email server.

Relay Host * **1**

From Email * **2**

From Name * **3**

SMTP Authentication Required * **4**
 Yes
 No

Username * **5**

Password * **6**

Cancel **OK**

- 1** Specify the relay host.
- 2** Specify the domain of the email address you would like emails to come from. In *example@company.com*, it is *company*.
- 3** Specify the name of the email address you would like emails to come from. In *example@company.com*, it is *example*.

- 4 If SMTP authentication is required, you must add a username and password. If you select **no**, you do not see the step to add a username and password.
- 5 Specify the username.
- 5 Specify the password.

Click **OK**.

Configure an email to receive alerts

ThoughtSpot sends alerts to the email address specified during installation. If you do not specify an email address, you do not receive any alerts. To add an email to receive alerts, `ssh` into your cluster from the command line and issue the following command.

Note: Add the ThoughtSpot Support alert email, `prod-alerts@thoughtspot.com`, to allow ThoughtSpot Support to receive alerts. ThoughtSpot Support monitors these alerts to ensure your cluster's health. Do not add this email to POC or demo environments.

```
$ tscli monitoring set-config --email <prod-alerts@thoughtspot.com>,<your_email>
```

To send to multiple emails, provide a comma-separated list with no spaces.

Verify the relay with an email

Check if the email settings are working properly by using this procedure.

1. Log in to the Linux shell using SSH.
2. Try sending an email to yourself by issuing:

```
$ echo | mail -s Hello <your_email>
```

3. If you receive the email at the address(es) you supplied, email is working correctly.

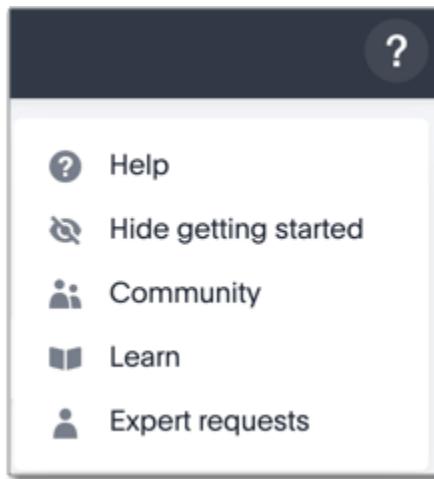
Customize ThoughtSpot Help

Summary: You can customize ThoughtSpot Help to be specific to your data, examples, and documentation.

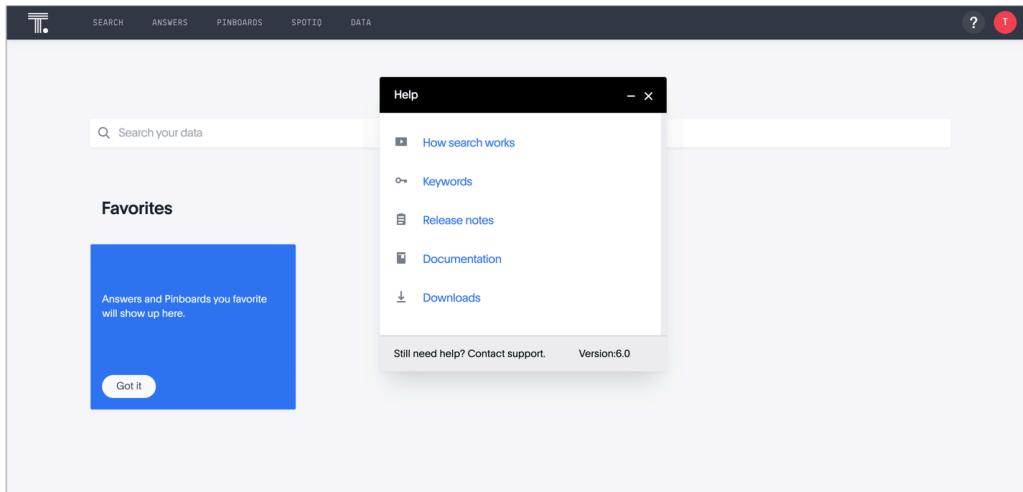
You can customize the Help menu for your ThoughtSpot application to tailor it to your organization.

When you configure these Help settings, you set system-wide defaults for all your users.

When your ThoughtSpot users click the Help icon, they see a list of links.

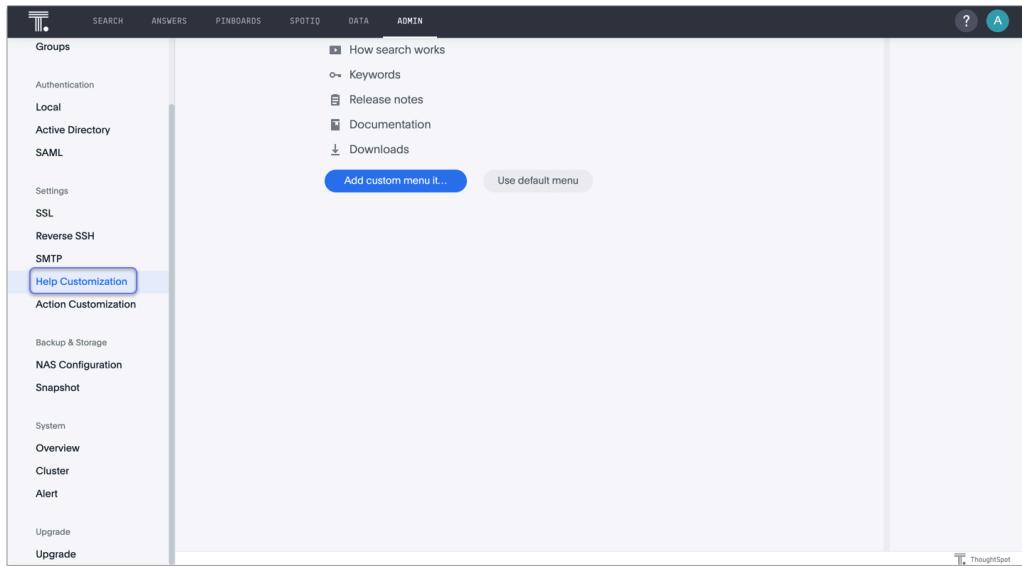


If a user clicks on the **help** link, the customizable help menu opens:

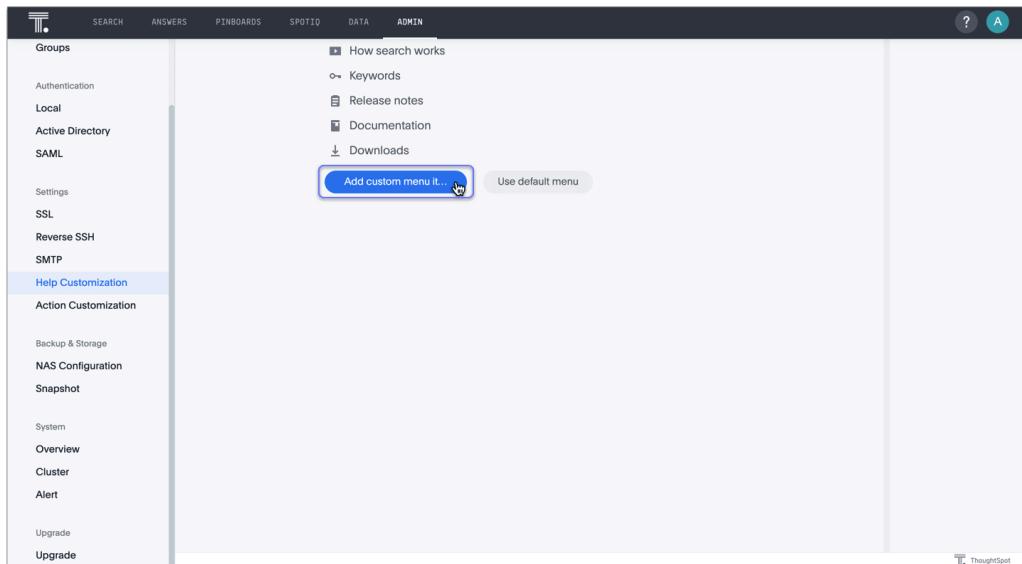


As an administrator, you can add your own links to this default list. This allows you to include documentation specific to your company, such as information about the data available in ThoughtSpot, where to get support internally, or company-specific training.

To customize the Help menu, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Help Customization** from the side navigation bar that appears.



To add a new link, select **Add custom menu item** below the list of existing links.



Specify the name of the link, the URL, and optionally add a custom icon. Then click **Confirm**.

Help menu customization

Item Label

URL

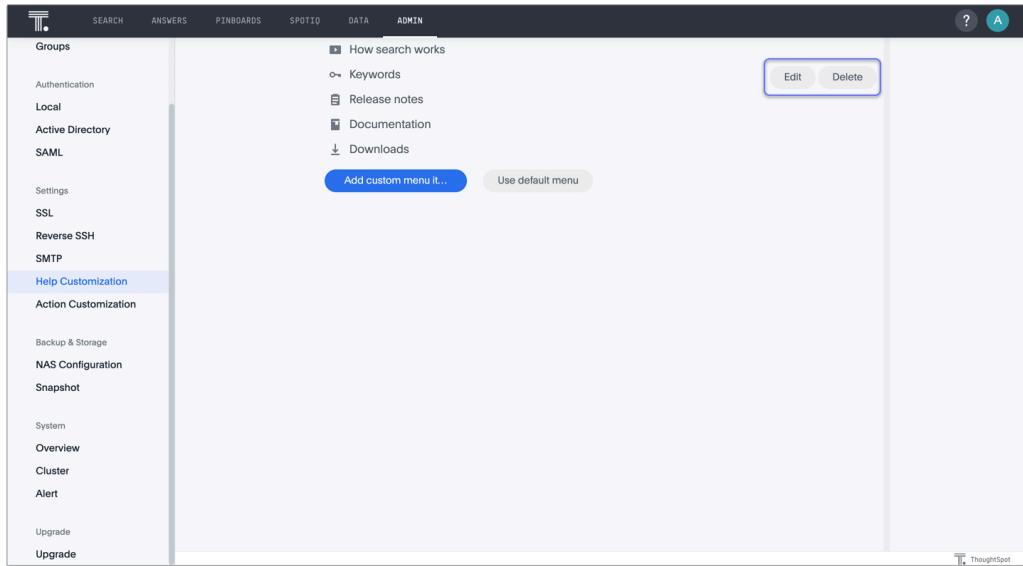
Custom Icon 

Item Enabled

Cancel **Confirm**

You can also edit existing links, change icons, or remove items from the help menu altogether.

To edit or delete a link, hover over the menu item and click the **Edit** or **Delete** button.



The screenshot shows the ThoughtSpot Admin interface with the 'Help Customization' section selected in the sidebar. The main area lists several help menu items:

- How search works
 - Keywords
 - Release notes
 - Documentation
- Downloads

Buttons for 'Edit' and 'Delete' are visible next to the 'How search works' item. A 'Use default menu' button is also present. The bottom right corner of the interface has the ThoughtSpot logo.

Specify the new information, and click **Confirm**.

Help menu customization

Item Label

URL

Custom Icon 

Item Enabled

Cancel Confirm

Customize Answer actions menu

Summary: You can add custom actions to the more options menu for Answers.

You can use the Data Push API and a custom menu item to initiate an action in another application based on the result of a search in ThoughtSpot.

These custom menu items are visible when you select the more options menu  on an Answer, or during a search, if you use the search terms that the external application expects to receive.

About the Data Push API

The Data Push API allows you to open a web page in the context of the ThoughtSpot application. This third party web page then has access to the ThoughtSpot search from which it was invoked. This is useful when you want to initiate an action in another application based on the result of a search in ThoughtSpot.

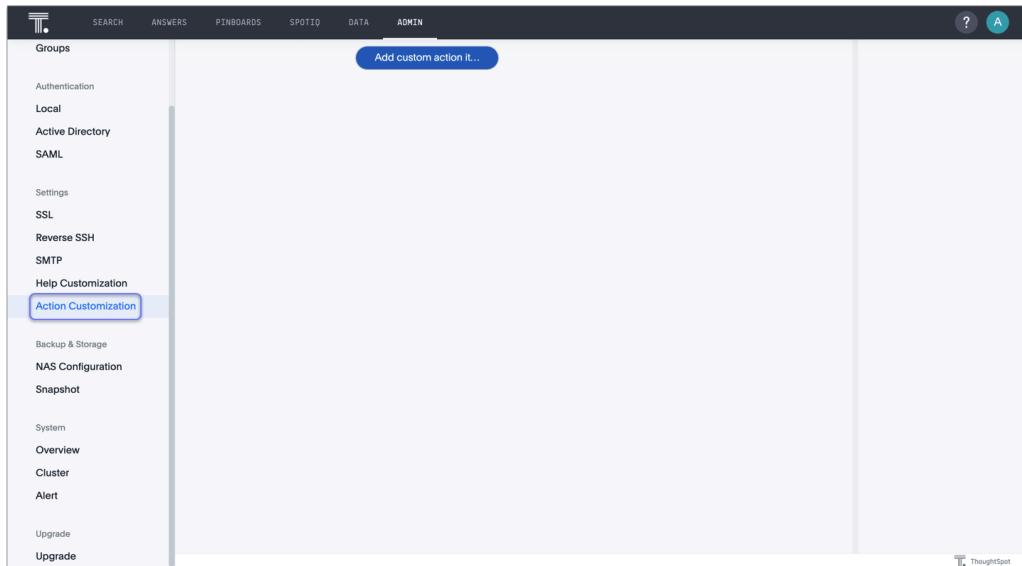
For example, you may search in ThoughtSpot for customers who are coming due for renewal of their contract in the next month. You could then trigger an action that brings up a web page from an external billing system. The billing system could be set up to read the data (list of names, emails, products, and renewal dates) from ThoughtSpot. The billing system might then add the price, generate an invoice for each customer, and send it by email.

To turn on the Data Push API functionality, [contact ThoughtSpot Support \[See page 0\]](#).

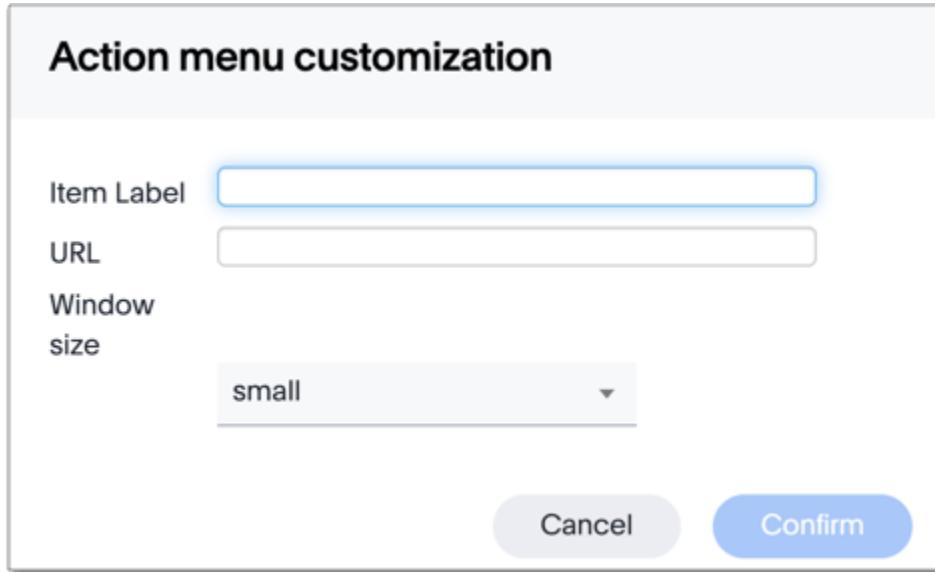
The data is available to the external application in JSON (JavaScript Object Notation) format. You must parse the JSON to get the data values you need using JavaScript in the receiving application.

Create a custom menu item from the Admin Console

To add custom items to the more options menu  for Answers, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Action Customization** from the side navigation bar that appears.



Select **Add custom action item**.



Specify a label for the item, such as *billing renewal email*, and add the URL of the external application.

Choose the size of the window you would like to open up in ThoughtSpot when a user clicks on this action, and click **Confirm**.

You can now see your action on ThoughtSpot searches and Answers, if you use the search terms that the external application expects to receive. Your action has a **custom** tag to differentiate it from ThoughtSpot actions.

When a user clicks on your action, the application you specified appears, and completes the actions specified by the Data Push API.

❶ Note: In order for your action to work correctly, you **must** use the search terms that the external application expects to receive in your Answer or unsaved ThoughtSpot search.

Style customization

Summary: Style customization allows you to change the overall style of your ThoughtSpot interface.

Using style customization, you can create a uniform ThoughtSpot experience that matches with your company's look and feel. To re-brand the interface, you can use the style customization option found in the Admin Console in the ThoughtSpot web application. It lets you change the logo, application background color, chart color palettes, and footer text. For help with chart and table visualization fonts, contact ThoughtSpot support.

This is especially useful if you're using the ThoughtSpot APIs for embedding visualizations from ThoughtSpot in your own web portal or application. You can make the visualizations match the look and feel of the portal or application in which they are embedded. For more information on using the APIs, see the ThoughtSpot Application Integration Guide.

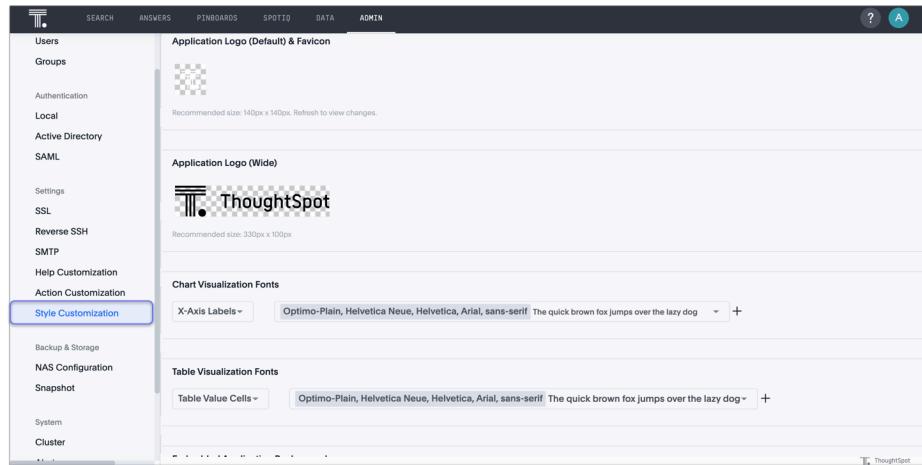
Style customization is not enabled by default. To enable style customization, [contact ThoughtSpot Support \[See page 0\]](#). The ThoughtSpot logo in the middle of the page is automatically removed when Style Customization is enabled.

Change style customization

Make changes to the style of your ThoughtSpot interface in the **Style Customization** page. This option gives you defined, yet impactful capabilities for re-branding the interface, so having some understanding of typography and color schemes would be helpful.

To re-brand the interface:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.
Select **Style Customization** from the side navigation bar that appears.



In the **Style Customization** page, you can perform the following actions:

- [Upload application logos \[See page 0\]](#)
- [Set chart and table visualization fonts \[See page 0\]](#)
- [Choose a background color \[See page 0\]](#)
- [Select chart color palettes \[See page 0\]](#)
- [Change the footer text \[See page 0\]](#)

Configure NAS for backup storage

Summary: You can use network attached storage (NAS) to support backup/restore and data loading.

About NAS mount

ThoughtSpot enables you to mount a NAS file system for storing or accessing large files. The file system mounts at the same location on each node in the cluster automatically. When any node restarts, the file system mounts again automatically, if it can be found.

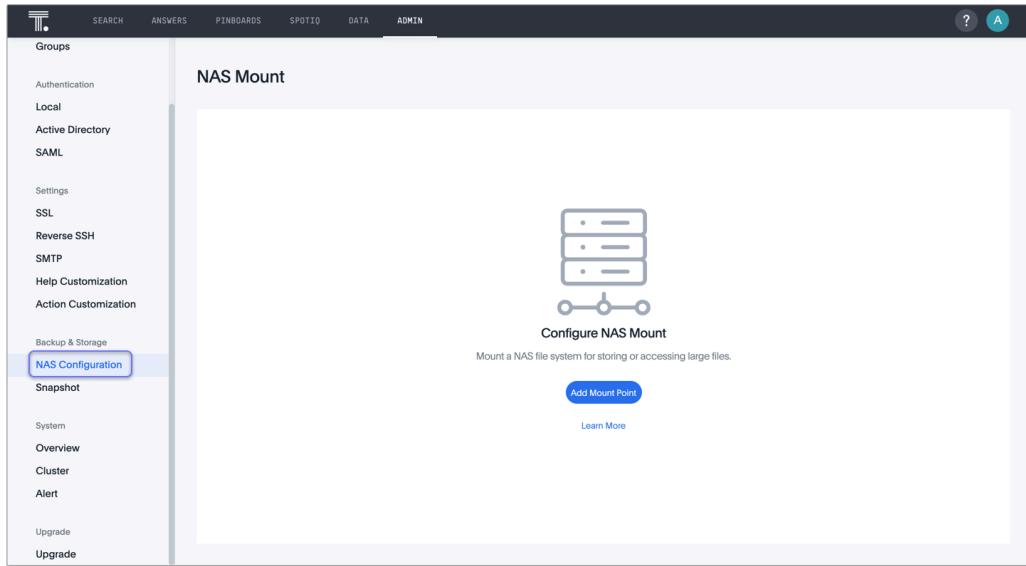
When supplying a directory for writing or reading a backup, you can specify a new mount point within `/export` as the directory to use. Likewise, you can stage data there for loading. It is best to have 2 separate NAS volumes, individually dedicated to data loads and backups.

Backups are written by the Linux user `admin`. If that user does not have permission to write to the NAS file system, you can write the backups to a disk (for example `/export/sdc1`, `/export/sdd1`, `/export/sde1`, or `/export/sdf1`) and then set up a cron job that executes as root user and copies the backup to the NAS device every night, then deletes it from the directory.

Do not send the periodic backups or stage files on `/export/sdb1` since it is a name node. It is used internally by Hadoop Distributed File System (HDFS) and if this drive fills up, it can cause serious problems. Do not allow backups or data files to accumulate on ThoughtSpot. If disk space becomes limited, the system will not function normally.

Mount NAS using the Admin Console

To mount a NAS file system using the Admin Console, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **NAS Configuration** from the side navigation bar that appears.



1. Select **Add Mount Point**.
2. Specify the parameters in the dialog box that appears:

Add Mount Point

Mount Type *

NFS 1

CIFS

Server Address * 2

Path on Server * 3

Local Mount Point * 4

Username * 5

Password * 6

Optional Mount Parameters

Other command-line options are available to forward to the command. 7

Cancel OK

- 1 Choose the mount type: either `NFS` (Network File System) or `CIFS` (Common Internet File System).
- 2 Specify the server address, in the form `storageservername.file.yourdomain.net`.
- 3 Specify the server path.
- 4 Specify the local mount point. This should be within the `/export` directory.
- 5 If you choose the `CIFS` mount type, you must specify a username and password.
- 6 If you choose the `CIFS` mount type, you must specify a username and password.

- 7 Optionally specify additional command-line subcommands and flags. Refer to the tscli reference [See page 0] for NAS commands.

3. Click **OK**.

You can unmount NAS from the Admin Console, or by running `tscli nas unmount --dir <directory>`.

Manage and create snapshots

Summary: A snapshot is a point-in-time image of your running cluster. You can use a snapshot to restore the cluster to a specific point in time. Create and manage snapshots from the Admin Console.

About manual snapshots

You must [create a snapshot \[See page 66\]](#) before making any changes to your cluster's environment, loading a large amount of data, or changing the structure of a table. ThoughtSpot supports up to 20 manual snapshots. You must clear them to create new snapshots. You can delete a snapshot [from the Admin Console \[See page 67\]](#).

ThoughtSpot generates a snapshot in approximately 20 seconds, depending on the size of the cluster. To restore from a snapshot, contact [ThoughtSpot Support \[See page 0\]](#).

❶ Note: During an upgrade, all snapshots from the previous version of ThoughtSpot become manual snapshots.

Manage snapshots through the Admin Console

From the Admin Console, you can [create manual snapshots \[See page 66\]](#), [delete existing snapshots \[See page 67\]](#), and view the details for all your existing snapshots.

To manage or create snapshots using the Admin Console, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Snapshot** from the side navigation bar that appears.

Snapshot Name	Scheduled On	Snapshot Style	Status
_x_periodic_20200720T140000_70640 2	20/07/2020 07:00	PERIODIC	Completed
_x_periodic_20200721T200000_38647 0	21/07/2020 13:00	PERIODIC	Completed
_x_periodic_20200722T100000_138090 8	22/07/2020 04:00	PERIODIC	Completed
_x_periodic_20200722T150000_138753 8	22/07/2020 08:00	PERIODIC	Completed
_x_periodic_20200722T190000_13960 8	22/07/2020 12:00	PERIODIC	Completed
_x_periodic_20200722T230000_140311 5	22/07/2020 16:00	PERIODIC	Completed
_x_periodic_20200723T030000_141146 5	22/07/2020 20:00	PERIODIC	Completed
_x_periodic_20200723T050000_141538 5	22/07/2020 22:00	PERIODIC	Completed
_x_periodic_20200723T190000_46205 5	23/07/2020 12:00	PERIODIC	Completed

You can view snapshot details, such as **name**, **style** (manual or periodic), **status**, and the **date and time** a snapshot was scheduled.

Create manual snapshots in the Admin Console

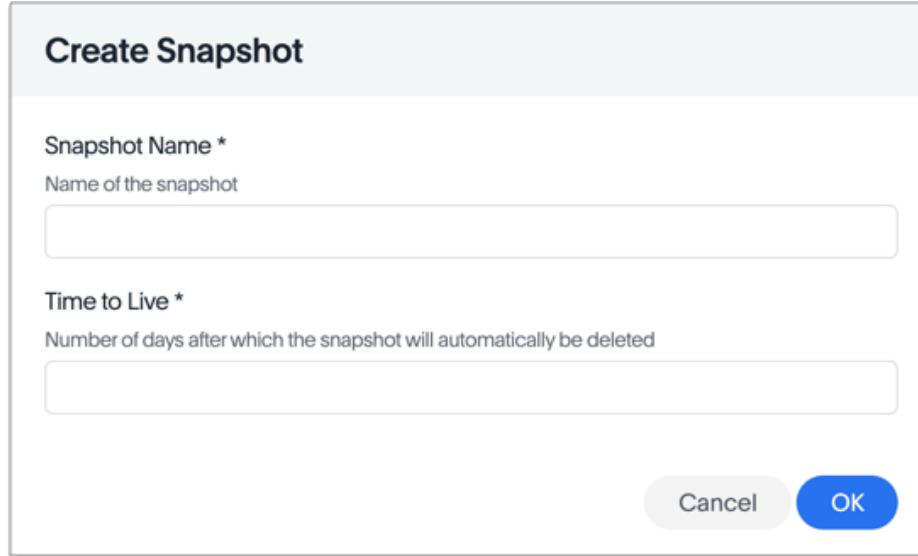
To create a manual snapshot in the Admin Console, follow these steps.

1. Click **Create** at the upper left side of your screen.

Snapshot Name	Scheduled On	Snapshot Style	Status
_x_periodic_20200720T140000_70640 2	20/07/2020 07:00	PERIODIC	Completed
_x_periodic_20200721T200000_38647 0	21/07/2020 13:00	PERIODIC	Completed
_x_periodic_20200722T100000_138090 8	22/07/2020 04:00	PERIODIC	Completed
_x_periodic_20200722T150000_138753 8	22/07/2020 08:00	PERIODIC	Completed
_x_periodic_20200722T190000_13960 8	22/07/2020 12:00	PERIODIC	Completed
_x_periodic_20200722T230000_140311 5	22/07/2020 16:00	PERIODIC	Completed
_x_periodic_20200723T030000_141146 5	22/07/2020 20:00	PERIODIC	Completed
_x_periodic_20200723T050000_141538 5	22/07/2020 22:00	PERIODIC	Completed
_x_periodic_20200723T190000_46205 5	23/07/2020 12:00	PERIODIC	Completed

2. Specify a **name** and a **time to live** for the snapshot.

If you do not want the system to delete the snapshot automatically after a specified number of days, enter `-1` as the **time to live**.



3. Click **OK**.

Delete snapshots in the Admin Console

To delete a snapshot in the Admin Console, follow these steps.

1. Click on the snapshot you would like to delete.
2. Select the **more options** menu icon  next to the snapshot name.
3. Click **remove**.

Manage and create snapshots

The screenshot shows the ThoughtSpot Admin interface. The left sidebar has a tree view with categories like User Management, Authentication, Settings, and Backup & Storage. Under Backup & Storage, the 'Snapshot' node is selected. The main panel displays a 'Snapshot Details' card for a completed snapshot. The card shows the following information:

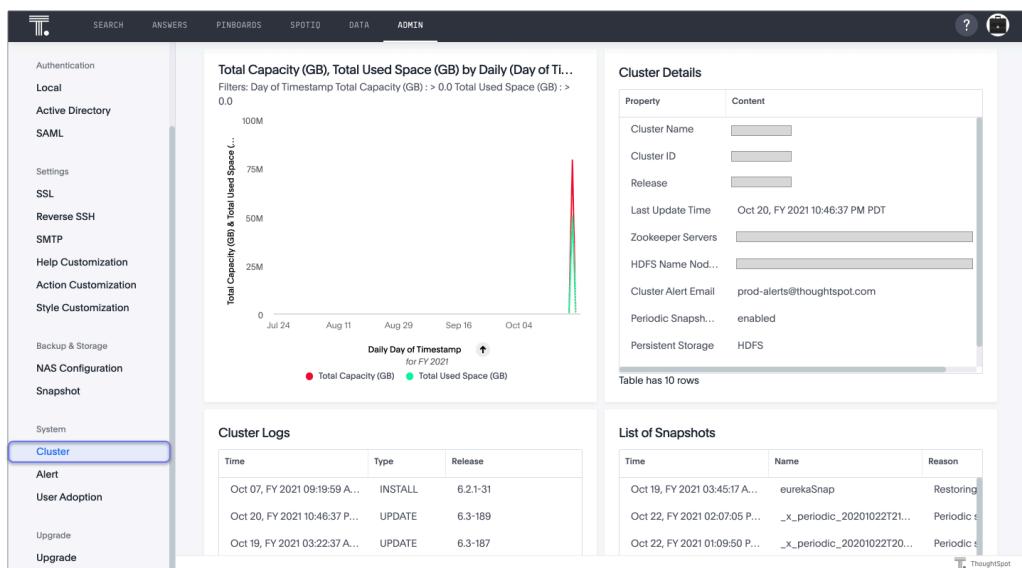
Snapshot Details	
Snapshot Name	_x_periodic_20200722T150000_138753
Type	PERIODIC
Scheduled On	22/07/2020 08:00
Status	Finished

A blue 'Remove' button is located in the top right corner of the card.

System Cluster Pinboard

Summary: View the System Cluster Pinboard in the Admin Console.

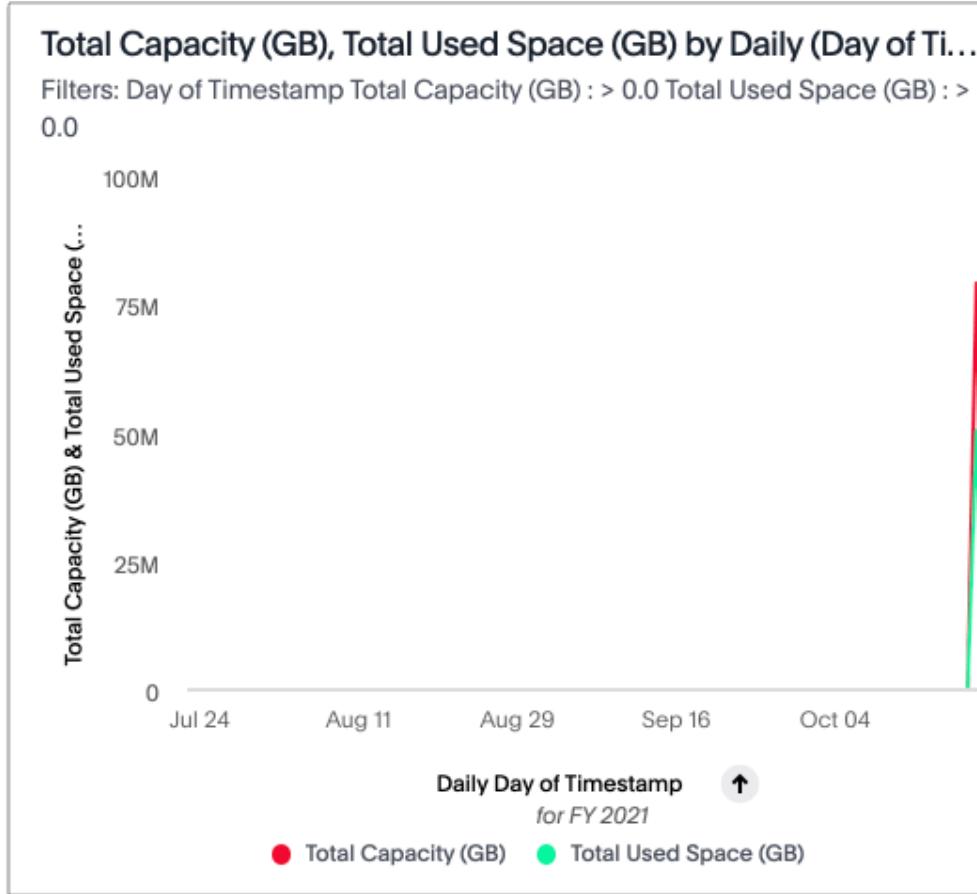
You can view cluster information from the Admin Console. The **System Cluster Pinboard** contains several ThoughtSpot Answers that display capacity, latency over time, snapshot status, installed release, node functions, and logs. To view this Pinboard, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Cluster** from the side navigation bar that appears.



Only the **Average Latency Last 2 Weeks (sec)** and **Total Capacity** panels contain a standard ThoughtSpot visualization. The other visualizations rely on internal queries that are not accessible. You can, however, **Present** them, or **Copy embed link**.

Total Capacity (GB), Total Used Space (GB) by Daily (Day of Timestamp)

The **Total Capacity (GB), Total Used Space (GB) by Daily (Day of Timestamp)** chart displays your cluster's total capacity and estimated used capacity over time.



The x-axis is by time and the y-axis measures the size in GB. You can zoom in and see daily or hourly utilization data. So, in the **Space Utilization** chart above, the green line shows the amount of capacity in use in the system, while the red line shows the total capacity. An increase in the red line at the end of a time period indicates the addition of extra hardware, resulting in increased capacity.

The query for this Answer is the following:

```
day of timestamp
total capacity (gb)
total used space (gb)
daily
last 90 days last 25 hours
total capacity (gb) > 0
total used space (gb) > 0
```

The chart relies on the `TS: Internal Table Wise Capacity WS` worksheet. It tracks total used space, which consists of raw uncompressed data, including replication.

Cluster Details

The fields on this visualization have the following meaning:

Field	Description
Cluster Name	The name of the cluster defined at installation time.
Cluster ID	The ID of the cluster set at installation time.
Last Update Time	Last time the cluster was updated.
Release	Version of the current release.
Zookeeper Servers	IP addresses of the Zookeeper servers.
HDFS Name Nodes	Control nodes for Hadoop Distributed File System (HDFS).

Cluster Logs

The fields on this visualization have the following meaning:

Field	Description
Time	A timestamp indicating when an action occurred.
Type	Type of action.
Release	Identifies the full release number.

You can also use the `tscli logs` command to review log data from your cluster.

List of Snapshots

This visualization shows the snapshots *and the backups* taken on the cluster. The fields on this visualization have the following meaning:

Field	Description
Time	A timestamp indicating when a snapshot or backup happened occurred.
Name	Name of the snapshot file. These files are stored in the `/usr/local/scaligent/backup` directory on your cluster.
Reason	Identifies the reason the snapshot/backup was created. You should see several period snapshots if your cluster is configured properly. You may also see evidence here of manual backups. For example, you should be sure your cluster is backed up before major events such as upgrades. Contact ThoughtSpot Support [See page 0] if you don't see evidence your cluster is periodically creating snapshots.
Size	Size of the backup in gigabytes.

Average Latency Last 2 Weeks (sec)

This visualization relies on the `TS: BI Server` worksheet to display the average database latency over the last 15 days. The database latency measures how long it takes for a search to return data from ThoughtSpot - this does not include the time taken to send the answer back to the client, it measures internal processing time. You can use the visualization menu to drill down to its underlying query:

```
average datacache (sec)
average total (sec)
daily
last 15 days
for database latency (us) > 0
```

System Alerts Pinboard

Summary: View the System Alerts Pinboard in the Admin Console.

You can view alert information for your ThoughtSpot cluster from the Admin Console. The **System Alert Pinboard** contains information on alerts, configuration events, and notification events. To view this Pinboard, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Alert** from the side navigation bar that appears.

Time	Type	Message
Jul 23, 2020 09:10:06 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:09:52 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:09:38 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:09:25 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:09:11 PM U...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:08:58 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:08:44 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:08:30 PM ...	TASK_TERMINAT...	Task falcon.worker to
Jul 23, 2020 09:07:57 PM ...	TASK_TERMINAT...	Task falcon.worker to

Table has 713 rows

Time	User	Summary
Jul 23, 2020 08:59:24 PM ...	tsadmin	User Management: I
Jul 23, 2020 08:58:11 PM U...	admin	tscli service add-jav
Jul 23, 2020 08:57:48 PM ...	admin	tscli service add-jav
Jul 23, 2020 08:40:23 PM ...	tsadmin	User Management: I
Jul 23, 2020 08:40:08 PM ...	tsadmin	User Management: I
Jul 23, 2020 08:38:23 PM ...	admin	tscli service add-jav
Jul 23, 2020 12:48:20 AM ...	tsadmin	User Management: I
Jul 23, 2020 12:47:41 AM U...	tsadmin	User Management: I
Jul 23, 2020 12:45:46 AM ...	paul@thoughtsp...	User Management: I

Table has 72 rows

Time	User	Summary

This Pinboard contains three ThoughtSpot Answers: **alerts**, **configuration events**, and **notification events**.

Alerts

The fields on this Answer have the following meaning:

Field	Description
Time	When the alert was sent.
Type	The ID of the event.

Message	The text of the alert message.
----------------	--------------------------------

For a full reference of possible alerts, see the [Alert code reference \[See page 0\]](#).

Configuration Events

This system answer displays recent events that changed the configuration of the system. This answer displays the **Time**, the **User** that performed the action, and a **Summary** of the action.

Notification events

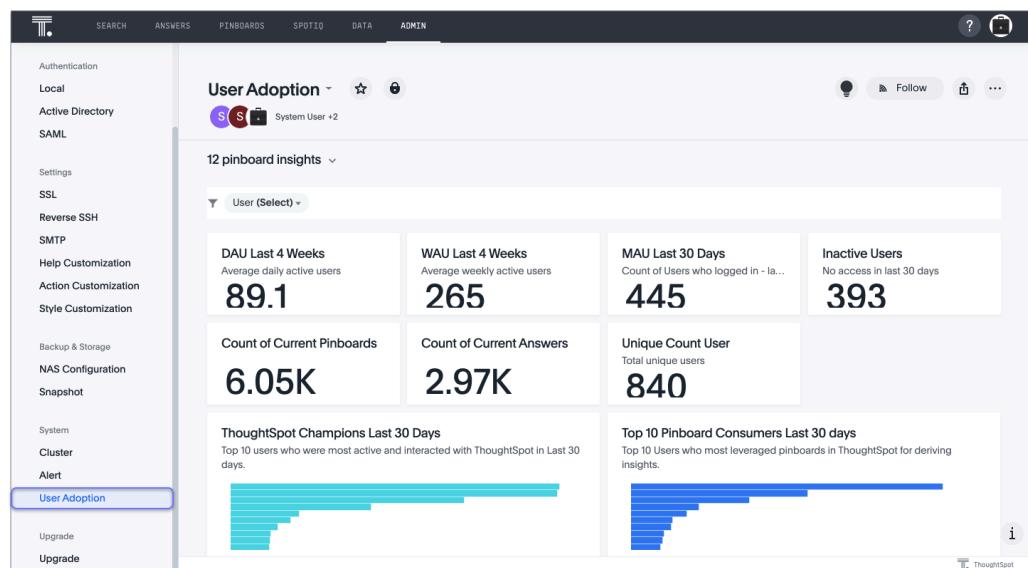
This answer displays notifications of data loads. The answer displays the **Time**, the **User** that performed the action, and a **Summary** of the action. Notifications are kept for 90 days before being discarded.

User Adoption Pinboard

Summary: Use the User Adoption Pinboard to understand how your ThoughtSpot users are interacting with ThoughtSpot, and how your user adoption is changing over time.

The User Adoption Pinboard contains essential information about your ThoughtSpot users and their adoption of the platform. Use this Pinboard to understand weekly active users, inactive users, top 10 adhoc searchers, popular Pinboards, and so on.

To view this Pinboard, navigate to the **Admin Console** by selecting **Admin** from the top navigation bar. Select **User Adoption** from the side navigation bar.



You can filter the Pinboard by user from the filter bar underneath the Pinboard name.

Use this Pinboard to view and analyze information about the following topics:

- hourly, daily, weekly, and monthly active users
- number of Pinboards and Answers on the cluster
- which users are most active in ThoughtSpot, and what they do
- popular Pinboards
- user actions in the last month
- unused Answers and Pinboards

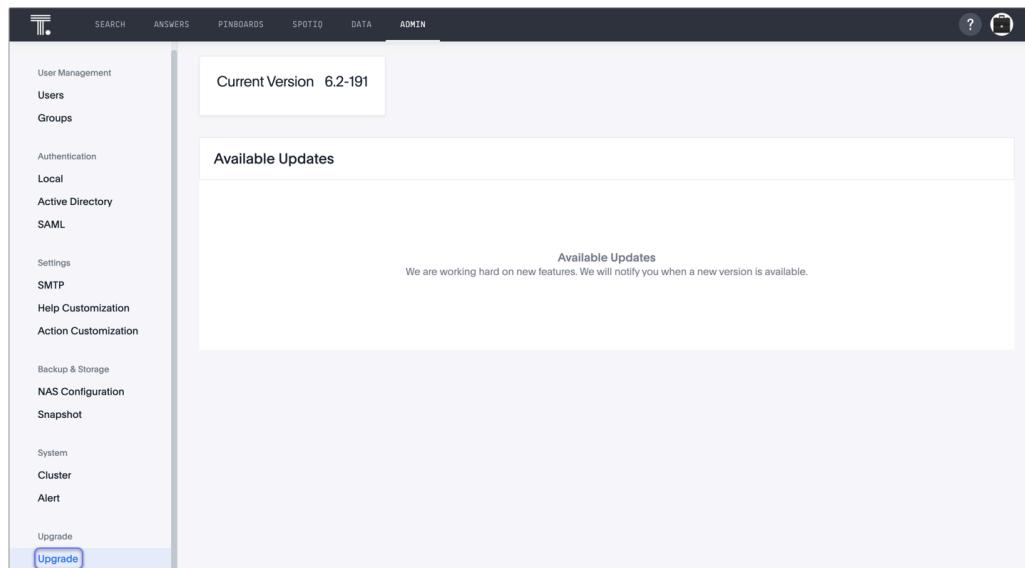
- inactive users
- users by last login, number of Pinboard interactions, number of Search interactions, and number of days logged in during the last 6 months

Available cluster updates

Summary: View available updates for your ThoughtSpot cluster.

You can view available updates for your ThoughtSpot cluster from the Admin Console. To view new releases, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Upgrade** from the side navigation bar that appears.

The **Upgrade** page shows your cluster's current version, and any new releases that are available. If no releases are available, you see the following screen.



If releases are available, and you would like to upgrade, [contact ThoughtSpot Support \[See page 0\]](#) to upgrade your cluster.

About installation and upgrades

Summary: As administrator, you are responsible for setting up and configuring ThoughtSpot. This guide explains how. It will also assist you in troubleshooting some common problems, finding additional resources, and contacting ThoughtSpot.

In ThoughtSpot, both physical appliances and virtual machines ship with a base OS image. ThoughtSpot Support then helps you to install, and subsequently update the software. Note the exception of RHEL installations, where the customer provides the OS image themselves, and uses a slightly different installation protocol.

The ThoughtSpot Support team contacts all clients to schedule updates as we release new versions of ThoughtSpot. To schedule upgrades to patch releases that support your use cases, work with our Customer Success team.

As administrator, you are responsible for setting up and configuring ThoughtSpot. This guide helps you with these tasks, and also assists in troubleshooting some common problems and finding additional resources.

Display your current configuration

To perform the setup and configuration, you must first learn how to [gain administrative access \[See page 8\]](#).

1. Log in to the ThoughtSpot cluster as the `admin` user.
2. Use the `tscli feature` subcommand to display your current configuration.

ACTION	NAME	STATUS	CONFIGUR
	Firewall	Disabled	
	Saml	Disabled	
	Ldap	Disabled	
	CustomBranding	Disabled	
	CustomBrandingFontCustomization	Disabled	
	DataConnect	Disabled	
	RLS	Enabled	
	Callhome	Enabled	
	SSHTunnel	Enabled	
	Fileserver	Disabled	

Related information

The following tasks are also available:

- [Set your locale \[See page 81\]](#)
- [Test connectivity between nodes \[See page 84\]](#)
- [Set up a fiscal calendar year \[See page 94\]](#)
- [Integrate LDAP \[See page 117\]](#)
- [Set up monitoring \[See page 137\]](#)
- [Configure support services \[See page 139\]](#)
- [Network ports \[See page 148\]](#)
- [Configure load balancing and proxies \[See page 160\]](#)

About installation and upgrades

- Customize look and feel [See page 166]

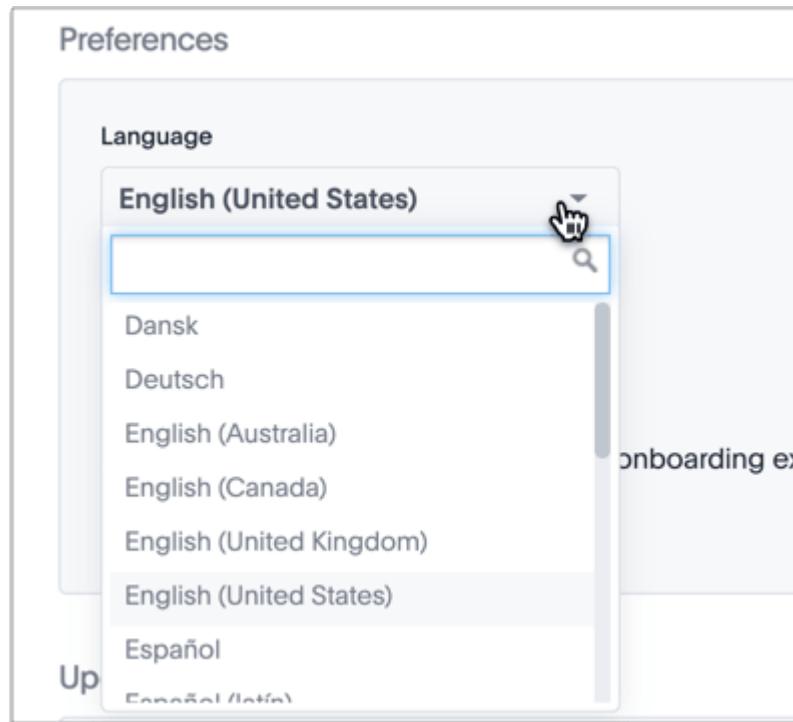
Set your ThoughtSpot locale

Summary: ThoughtSpot supports many locales. Learn how to set your display language, date formatting, and currency format settings by specifying your locale.

The language that ThoughtSpot displays, the date format, and the currency format depend on the system locale. If you set your browser locale or OS locale to one of ThoughtSpot's supported locales, ThoughtSpot automatically defaults to that locale. If your browser locale is not one of ThoughtSpot's supported locales, ThoughtSpot defaults to US English.

It is simple to change this locale using the **Profile** interface.

1. Navigate to your user profile:
 - Select your user icon next to the **help** button.
 - Select **Profile**.
2. Under **Preferences**, click the drop-down menu icon to change the language.



The **Language** selection specifies more than just the language: it sets the locale, which controls both the language choice and standard data formats for date and number. So, if you set French as the default locale in your profile settings, the interface updates to reflect this. Be sure to refresh your browser page.

For example, in the United States the number format for large numbers uses the comma thousands separator and a period decimals separator, and looks like this: `xxx,xxx.xx`. Most European countries use the reverse notation, with a comma decimals separator and a period thousands separator, like this:

`xxx.xxx,xx`.

In addition to American English (`en-US`), ThoughtSpot supports the following locales:

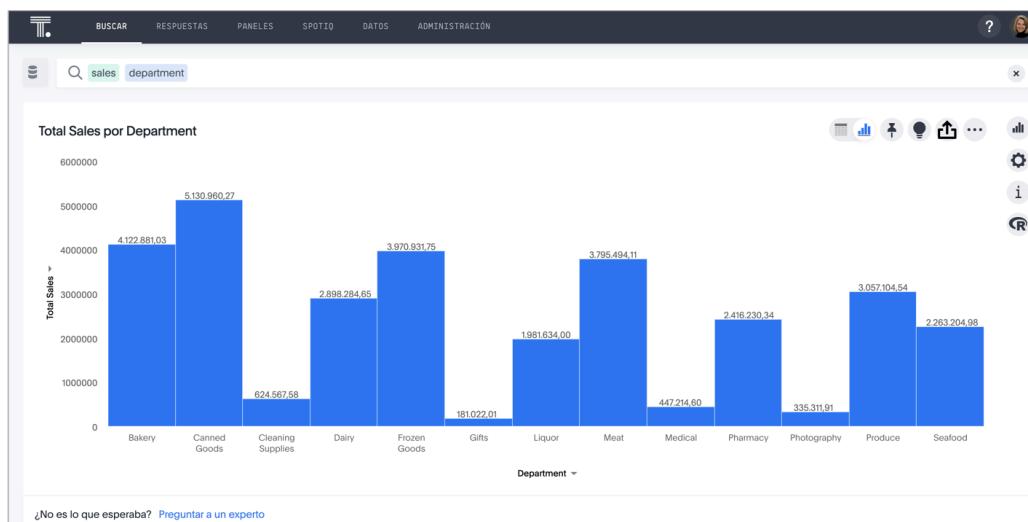
Locale	Language
<code>da-DK</code>	Dansk
<code>de-DE</code>	Deutsche
<code>en-AU</code>	English (Australia)
<code>en-CA</code>	English (Canada)
<code>en-GB</code>	English (United Kingdom)
<code>en-US</code>	English (United States)
<code>es-US</code>	Español (Latinoamérica)
<code>es-ES</code>	Español (España)
<code>fr-CA</code>	Français (Canada)
<code>fr-FR</code>	Français (France)
<code>it-IT</code>	Italiano
<code>nl-NL</code>	Nederland
<code>nb-NO</code>	Norsk
<code>pt-BR</code>	Português (Brasil)
<code>pt-PT</code>	Português (Portugal)
<code>fi-FI</code>	Suomi
<code>sv-SE</code>	Svenska

Locale	Language
zh-CN	中文(简体)
ja-JP	日本語

ThoughtSpot translates keywords, operators, and error messages. See the [keyword reference \[See page 0\]](#) for all supported languages.

ThoughtSpot *DOES NOT* translate formulas, or metadata entered by the user. For example, if you name a visualization ‘Quarterly Sales’ in any variant of English and subsequently change the locale to a variant of French, the visualization remains ‘Quarterly Sales’ and does not become ‘Ventes trimestrielles’. If you specify a currency when uploading data, that currency does not change when the locale changes.

Here is an example of a locale change from the default, American English, to Spanish.



Note that the top navigation bar is now in Spanish. The number formatting of the data markers now reflects the European format, with a comma decimal separator, and a period thousands separator. However, the column names are still in English, because ThoughtSpot does not translate your metadata.

Test network connectivity between nodes

Summary: Verify your network is properly configured for the application.

This procedure tests the network connectivity between all ThoughtSpot nodes, and to the LAN. If you can perform these steps successfully, the network settings on ThoughtSpot are correct.

1. Log in to the Linux shell using SSH.
2. Ping each of the other nodes in the cluster.
3. Ping another machine that exists outside of the cluster, for example, a machine that you will use to stage data to be loaded.

If you cannot perform these tests successfully, there is a problem with the network setup. If the tests fail, check [Network connectivity issues \[See page 632\]](#).

In-app acceptance of ThoughtSpot's use agreement

Summary: An administrator must sign the ThoughtSpot use agreement in the application to allow users to continue using the application.

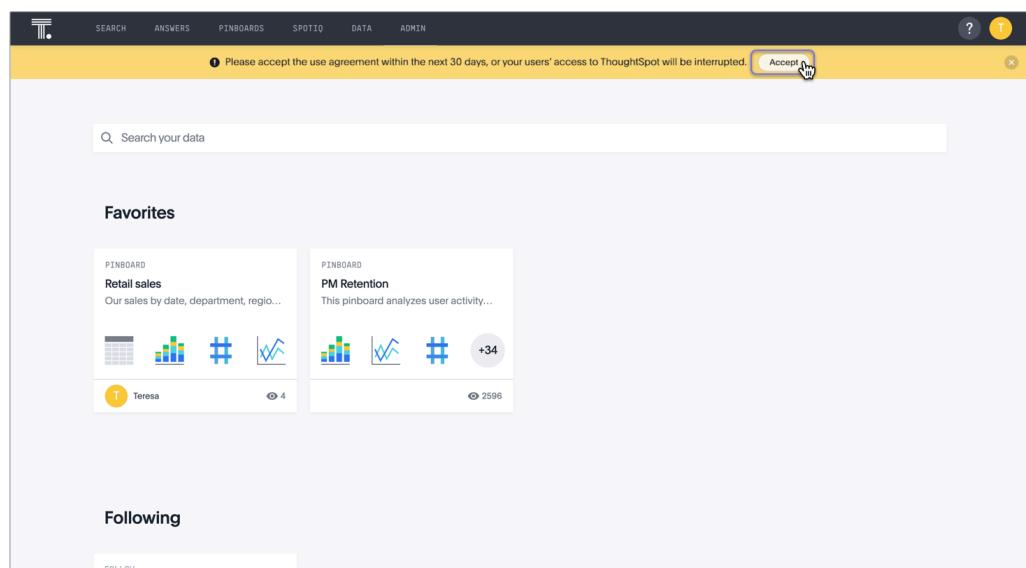
Prior to ThoughtSpot release 6.2, administrators signed a paper use agreement to allow their users to use ThoughtSpot. Starting with release 6.2, you can easily sign the use agreement from the application itself. You can view a copy of this agreement at [ThoughtSpot Legal](https://www.thoughtspot.com/legal) (<https://www.thoughtspot.com/legal>).

Note: Even if you previously signed a paper copy of the use agreement, an admin **must** sign the agreement in the application, within 30 days of your upgrade to release 6.2 or later.

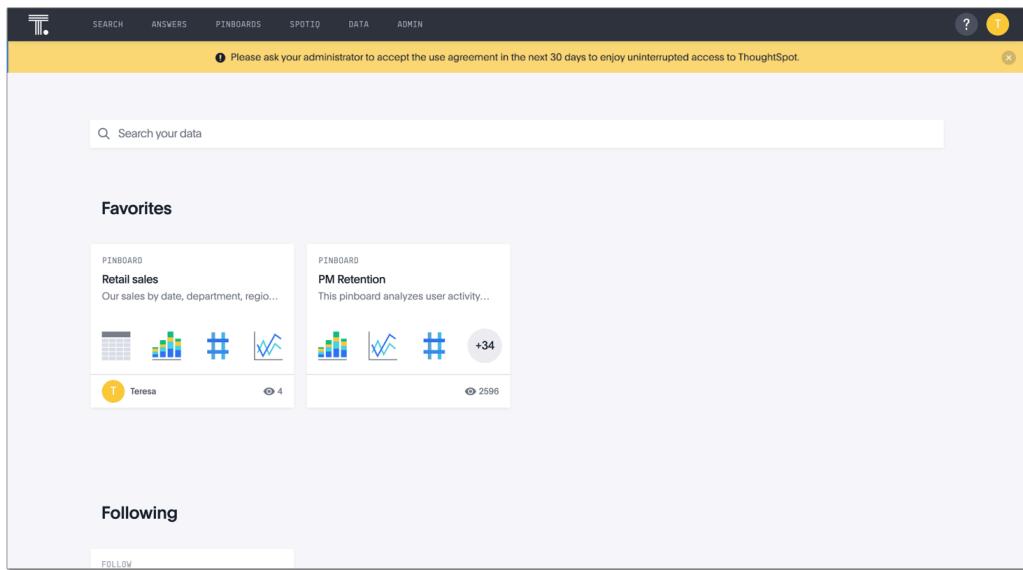
Use agreement notifications

When you upgrade your cluster to 6.2, or when there is a new version of the use agreement later, you receive notifications about the use agreement. You do not receive notifications about the use agreement with every software upgrade, just with upgrades that include changes to the use agreement.

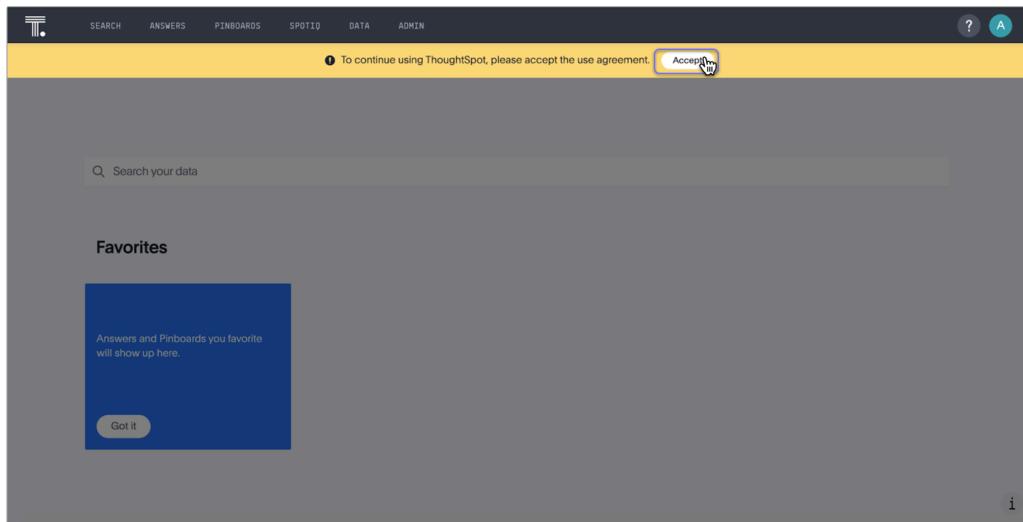
Administrators receive the following notification, which allows them to accept the use agreement:



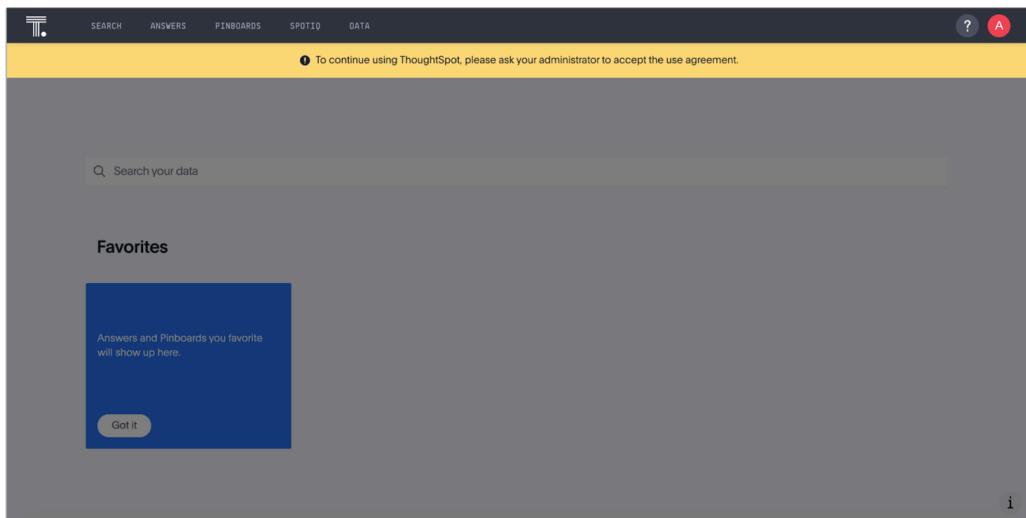
Users who do not have administrator privileges receive the following notification, which tells them to ask their administrator to accept the use agreement:



One administrator for each cluster must accept this agreement within 30 days of upgrading to version 6.2 or later. If an administrator does not sign the agreement within 30 days, users cannot access the ThoughtSpot application. Administrators receive the following notification when they try to use ThoughtSpot, which allows them to accept the use agreement to continue using ThoughtSpot:



If an administrator does not sign the agreement within 30 days of upgrading the cluster to 6.2, non-admin users receive the following notification when they try to use ThoughtSpot. The notification tells them to ask their administrator to accept the use agreement to allow them to continue using ThoughtSpot:

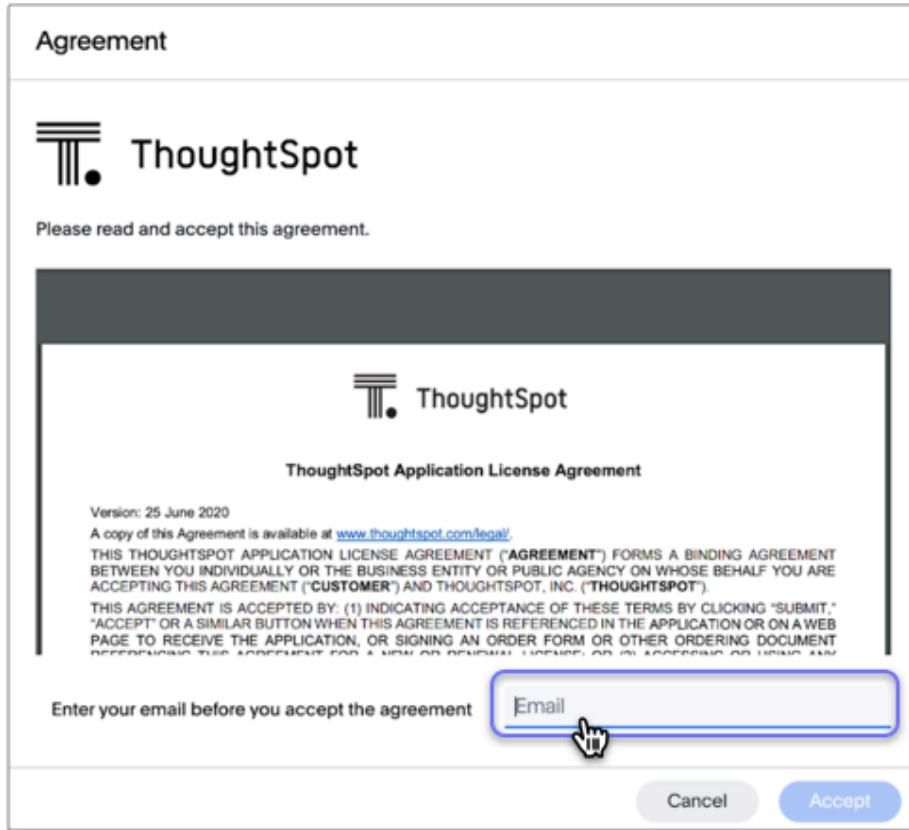


Accept the use agreement

Only administrators can accept the use agreement. Only one administrator needs to accept the use agreement for each cluster. Note that you must sign an agreement for each of your clusters.

To accept the use agreement, follow these steps. You do not need internet access to sign the agreement, since it is bundled with the application software.

1. Ensure that you have administrator privileges.
2. Select **Accept** from the yellow notification bar.
3. Read the use agreement in full before accepting. You can view a copy of the agreement at [ThoughtSpot Legal](https://www.thoughtspot.com/legal) (<https://www.thoughtspot.com/legal>).
4. If your ThoughtSpot account does not have an associated email, you must enter your email before you can sign the agreement. If your ThoughtSpot account has an associated email, you do not need to enter it in this window. The email text box does not appear if you already have an email associated with your account.



5. Click **Accept**.

If you have a technical issue when you sign the use agreement, and users cannot access the application, contact [ThoughtSpot Support](#) [See page 0].

Set the relay host for SMTP (email)

Summary: ThoughtSpot uses emails to send critical notifications to ThoughtSpot Support. A relay host for SMTP traffic routes the alert and notification emails coming from ThoughtSpot through an SMTP email server.

You can configure the relay host [using tscli](#) [See page 89] or through the Admin Console [See page 91].

Configure using tscli

Set up SMTP rules to send critical email notifications to ThoughtSpot Support.

Set up the relay Host

To set up a relay host:

1. Log in to the Linux shell using SSH.
2. Issue the setup command, providing the IP address of the relay host:

On ThoughtSpot release 6.1.1 and later, or on release 6.0.5, you can specify a custom port to connect to the relay host. If you do not specify a port, the system uses the default recommended port, port 25. Use a custom port if port 25 is blocked in your environment.

To use the default port, run the setup command:

```
$ tscli smtp set-relayhost <IP_address>
```

To use a custom port instead of port 25, run the setup command, specifying the port you want to use:

```
$ tscli smtp set-relayhost <IP_address>:<custom_port>
```

If you are on 6.1 rather than 6.1.1, or an earlier version than 6.0.5, [contact ThoughtSpot Support](#) [See page 0] to use a custom port.

3. Verify your settings:

```
$ tscli smtp show-relayhost
```

4. Verify that email is working.

Configure an email to receive alerts

ThoughtSpot sends alerts to the email address specified during installation. If you do not specify an email address, you do not receive any alerts. To add an email to receive alerts, issue the following command.

Note: Add the ThoughtSpot Support alert email, `prod-alerts@thoughtspot.com`, to allow ThoughtSpot Support to receive alerts. ThoughtSpot Support monitors these alerts to ensure your cluster's health. Do not add this email to POC or demo environments.

```
$ tscli monitoring set-config --email <prod-alerts@thoughtspot.com>,<your_email>
```

To send to multiple emails, provide a comma-separated list with no spaces.

Verify the relay with an email

Check if the email settings are working properly by using this procedure.

1. Log in to the Linux shell using SSH.
2. Try sending an email to yourself by issuing:

```
$ echo | mail -s Hello <your_email>
```

3. If you receive the email at the address(es) you supplied, email is working correctly.

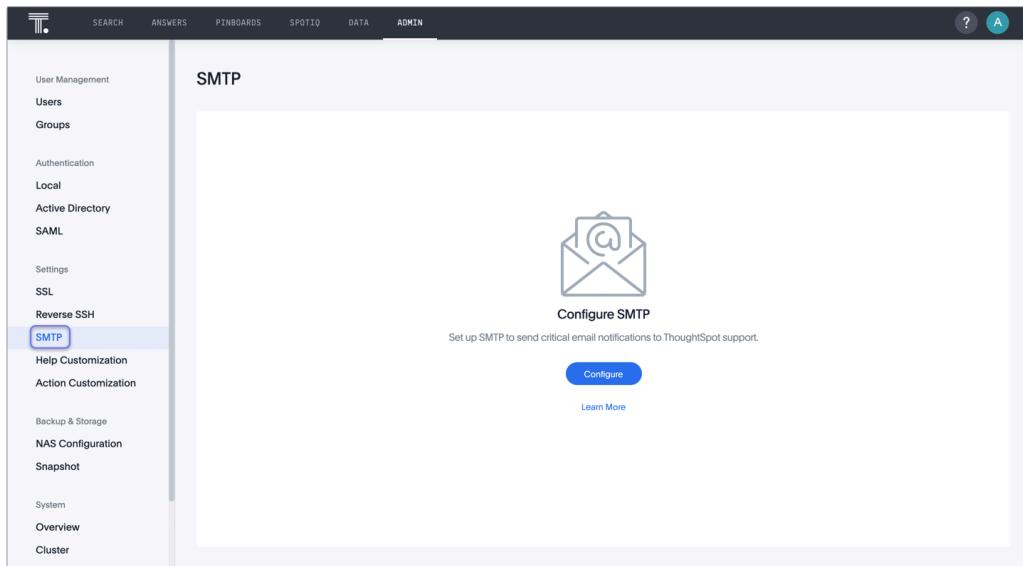
Configure SMTP through the Admin Console

You can set up the relay host for SMTP from the Admin Console.

Note: If you would like to use a custom port, rather than the default, port 25, you must configure SMTP using tscli, by running `tscli smtp set-relayhost <IP_address>:<custom_port>`.

Set up relay host

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SMTP** from the side navigation bar that appears.



Select **Configure**.

Specify the following parameters:

Configure SMTP

A relay host for SMTP traffic routes the alert and notification emails coming from ThoughtSpot through an SMTP email server.

Relay Host * 1

From Email * 2

From Name * 3

SMTP Authentication Required * 4

Yes
 No

Username * 5

Password * 6

Cancel OK

- 1 Specify the relay host.
- 2 Specify the domain of the email address you would like emails to come from. In `example@company.com`, it is *company*.
- 3 Specify the name of the email address you would like emails to come from. In `example@company.com`, it is *example*.
- 4 If SMTP authentication is required, you must add a username and password. If you select **no**, you do not see the step to add a username and password.

5 Specify the username.

5 Specify the password.

Click **OK**.

After you configure SMTP through the Admin Console, use `tscli` to [configure an email to receive monitoring alerts \[See page 90\]](#) and [verify the relay with an email \[See page 90\]](#).

Set up custom calendars

Summary: With a custom calendar, you can choose when the year, quarter, or week starts, and search using date-related keywords.

You can create custom calendars for different purposes within your company, including:

- Retail calendars, like 4-4-5, 4-5-4, or 5-4-4
- Fiscal calendars, where the year or quarter starts on a different date each year
- Pharmaceutical calendars, where the week runs from Friday through Thursday
- Calendars in different languages

Options

In a custom calendar, you can set the following:

- The starting date of a year
- The starting dates of quarters
- The starting day of the month
- The starting day of the week
- The words used for the days of the week, months and quarters

Search features

You can use your custom calendar to search the following ways:

- Using date keywords, like `this quarter` and `q3`
- Using date formulas with the `fiscal` option specified (See <https://docs.thoughtspot.com/6.0/advanced-search/formulas/date-formulas.html#fiscal-and-gregorian-calendars>)
- Overriding the calendar used in the search bar by typing your custom calendar's name

Limitations

- Maximum scope of the date dimension is 90 years
- Maximum length of a calendar year is 12 months

Setting up a custom calendar

To set up a custom calendar for your cluster, you must do the following:

1. Enable the custom calendar feature.
2. Generate a calendar template.
3. Edit the calendar template.
4. Add the custom calendar to your cluster.

Enable the custom calendar feature

To enable the custom calendar feature for your cluster, contact ThoughtSpot Support (See <https://docs.thoughtspot.com/latest/admin/misc/contact.html>).

Generate a calendar template

Using a calendar template as your starting point ensures that you use a format that is compatible with ThoughtSpot.

To generate a calendar template, do the following:

1. SSH as admin into your ThoughtSpot cluster: `ssh admin@<cluster-ip-address or hostname> .`
2. Run the `tscli calendar generate` command using the following syntax:

```
tscli calendar generate --name <calendar_name> --start_date <MM/DD/YYYY> --
end_date <MM/DD/YYYY> --username tsadmin
```

Example: `tscli calendar generate --name my_calendar --start_date 07/01/2019
--end_date 06/30/2020 --username tsadmin`

This generates a calendar template file in .csv format. In the previous example:
my_calendar.csv.

3. Exit your SSH session.

Edit the calendar template

To use the template you generated as your custom calendar, some editing is required.

1. Download the .csv file to your computer using following syntax:

```
scp admin@<cluster-ip-address>:/home/admin/<calendar_name>.csv /<Local  
directory on your computer>/.
```

Example (on Mac OS): `scp admin@172.18.144.217:/home/admin/my_calendar.csv
/Users/john.smith/Desktop/.`

2. Open the .csv file in a text editor or spreadsheet program and edit the file to ensure the date and quarter columns are formatted correctly:
 - The Date column must use the format: **MM/DD/YYYY**. No other formats are supported.
 - The Quarter column must display the correct quarter number for each day of the year.

❶ Note: By default, a generated calendar displays quarter numbers based on the Gregorian calendar (which starts on January 1st). If your custom calendar begins any other date, you must adjust the quarter numbers to align with your calendar. For example: If your custom calendar begins on April 1st, the calendar would incorrectly show April, May and June as quarter 2. You would need to correct this to indicate those months are quarter 1 and correct the subsequent months to have the correct quarter.

- (Optional) To enhance searchability, ThoughtSpot recommends adding a “Q” before each quarter number. Example: **Q1**. If adapting the calendar to different language, use the appropriate letter in place of “Q”.
- Make any other changes needed to the calendar (like translating months or days into a different language.)

Example calendar with the fiscal year beginning on April 1:

date	day_of_week	month	quarter	year	day_number_of_week	week_number_of_month	week_number_of_quarter	week_number_of_year	is_weekend
04/01/2019	monday	april	Q 1	2019	1	1	1	1	FALSE
04/02/2019	tuesday	april	Q 1	2019	2	1	1	1	FALSE
04/03/2019	wednesday	april	Q 1	2019	3	1	1	1	FALSE
04/04/2019	thursday	april	Q 1	2019	4	1	1	1	FALSE
04/05/2019	friday	april	Q 1	2019	5	1	1	1	FALSE
04/06/2019	saturday	april	Q 1	2019	6	1	1	1	TRUE
04/07/2019	sunday	april	Q 1	2019	7	2	1	1	TRUE
04/08/2019	monday	april	Q 1	2019	1	2	2	2	FALSE
04/09/2019	tuesday	april	Q 1	2019	2	2	2	2	FALSE
04/10/2019	wednesday	april	Q 1	2019	3	2	2	2	FALSE
04/11/2019	thursday	april	Q 1	2019	4	2	2	2	FALSE
04/12/2019	friday	april	Q 1	2019	5	2	2	2	FALSE
04/13/2019	saturday	april	Q 1	2019	6	2	2	2	TRUE
04/14/2019	sunday	april	Q 1	2019	7	3	2	2	TRUE
04/15/2019	monday	april	Q 1	2019	1	3	3	3	FALSE
04/16/2019	tuesday	april	Q 1	2019	2	3	3	3	FALSE
04/17/2019	wednesday	april	Q 1	2019	3	3	3	3	FALSE

- Save your calendar template as a UTF-encoded .csv file with UNIX line breaks.

Note: Saving the file with UNIX line breaks, ensures there are no carriage returns (^M characters) in the file which prevent you from using your calendar in ThoughtSpot. Microsoft Excel, for example, adds carriage returns. The easiest way to remove carriage returns is to open your .csv file in a text editor, and save it as a .csv with UNIX line breaks.

Add the custom calendar to your cluster

To use your edited calendar template as a custom calendar, you must upload it to your cluster and use it to create a calendar in ThoughtSpot.

- Upload the .csv file to your ThoughtSpot cluster using the following syntax:

```
scp /<Local directory on your computer>/<calendar_template_name>.csv
admin@<cluster-ip-address>:/home/admin/
```

Example (on Mac OS): `scp /Users/john.smith/Desktop/my_calendar.csv`
`admin@172.18.144.217:/home/admin`

- SSH as admin into your ThoughtSpot cluster: `ssh admin@<cluster-ip-address or hostname> .`

- Run the `tscli calendar create` command using the following syntax: `tscli calendar create --file_path /home/admin/<calendar_template_name>.csv --name <calendar name> --username tsadmin`

```
Example: tscli calendar create --file_path /home/admin/my_calendar.csv --name my_calendar --username tsadmin
```

(Optional) Set a custom calendar as the default calendar for your cluster

To set your custom calendar as the default calendar for your cluster, contact ThoughtSpot Support (See <https://docs.thoughtspot.com/latest/admin/misc/contact.html>).

Setting a worksheet, table or view to use your custom calendar

If you don't set your custom calendar as the default for your cluster, you must do the following to use your calendar:

1. Sign in to your ThoughtSpot cluster and click **DATA**.
2. On the DATA page, click the name of a worksheet, table or view in which you want to use your custom calendar.
3. Under COLUMN NAME, find a column that uses the DATE or DATE_TIME data type where you want to use your custom calendar and scroll right until you see the CALENDAR TYPE column.

Note: The column must use the DATE or DATE_TIME data type.

4. In the CALENDAR TYPE column for the column(s) you chose, double-click the existing calendar name, and then select your custom calendar.
5. Click **Save Changes**.

Now, date-related searches in the selected worksheet, table or view use your custom calendar.

Configure internal authentication

Summary: Many organizations don't use LDAP or active directory for user authentication. In other scenarios, some users are not in LDAP and are created only in ThoughtSpot. In such cases, you can choose to authenticate users against ThoughtSpot internal authentication.

Many organizations don't use LDAP or active directory for user authentication. In other scenarios, some users are not in LDAP and are created only in ThoughtSpot. In such cases, you can choose to authenticate users against ThoughtSpot internal authentication.

Note: Before the user logs into ThoughtSpot, the user must exist in ThoughtSpot. This is independent of the authentication mechanism.

Enable internal authentication with the Admin Console

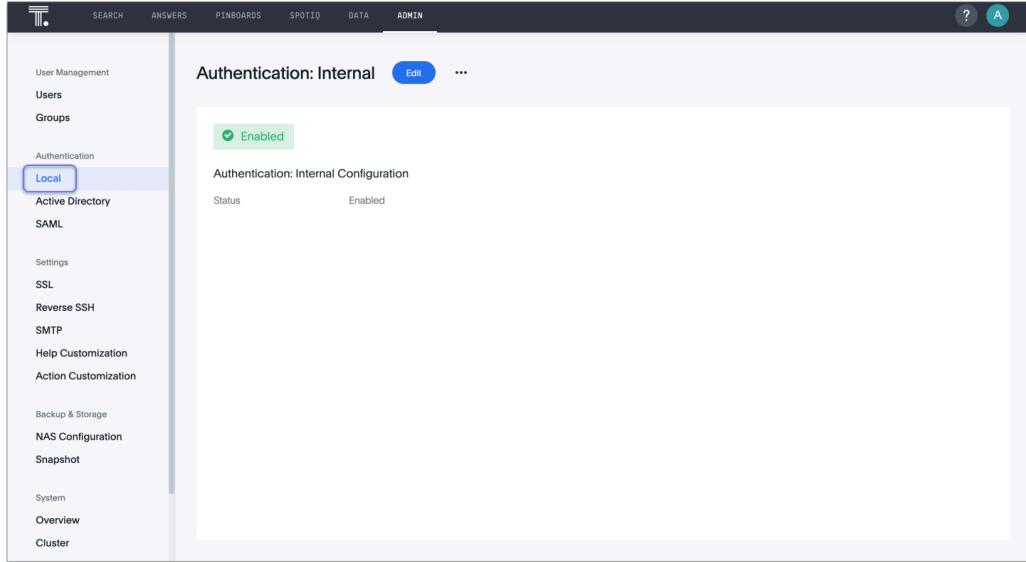
Local authentication is enabled by default in your ThoughtSpot environment. This means that any local user you [create \[See page 25\]](#) can log into ThoughtSpot with the specific ThoughtSpot credentials you set up for them when creating the user.

If you disable local authentication, no local user can log into ThoughtSpot.

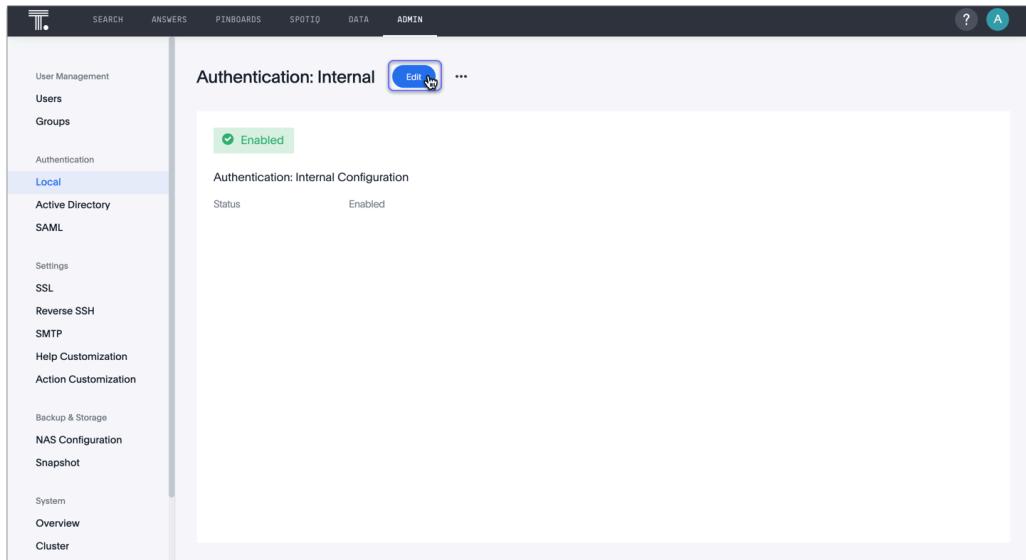
You can enable or disable local authentication from the Admin Console.

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Local** from the side navigation bar that appears.

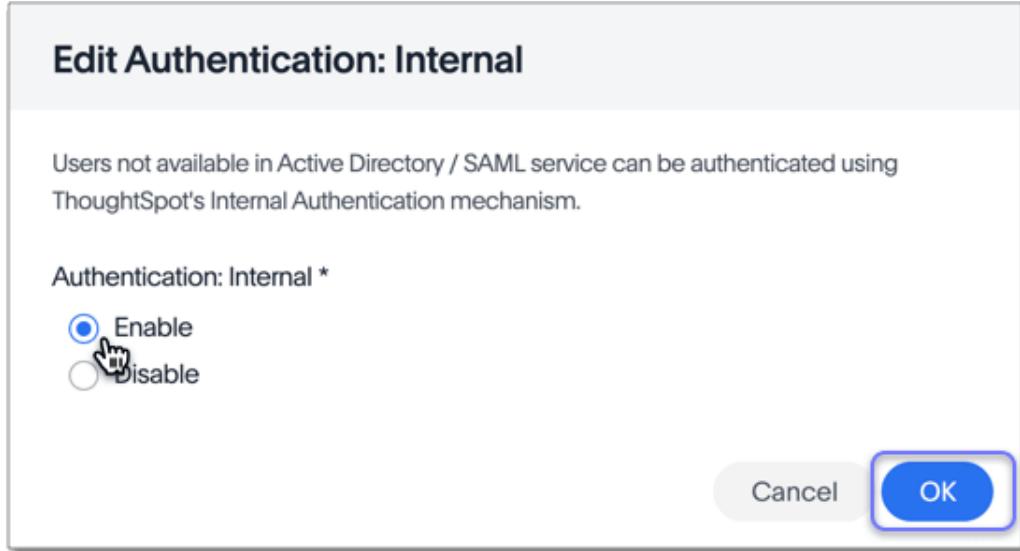
Configure internal authentication



Click the **Edit** button in the top middle of the screen.



Select **Enable** or **Disable**, and click **OK**.



Note that you cannot disable local authentication until you configure [Active Directory \[See page 34\]](#) or [SAML authentication \[See page 31\]](#).

If you have been using ThoughtSpot with users you created manually, and you now want to transition to LDAP, contact [ThoughtSpot Support \[See page 0\]](#). They can assist you in migrating existing users to their LDAP equivalents.

Configure SSL

Summary: Secure socket layers (SSL) provide authentication and data security when sending data to and from ThoughtSpot.

You can use your own SSL certificate to secure ThoughtSpot HTTP(S) traffic.

About SSL

To enable SSL for the ThoughtSpot web service, you must provide your organization's SSL certificate for the ThoughtSpot service URL. If you do not have this certificate, you have the following options:

- Check with your IT department to see if they have an SSL certificate you can use.
- Obtain the certificate from an issuing authority.
- Use the default SSL certificate on the ThoughtSpot nodes.
- Disable SSL using the `tscli ssl off` command.

ThoughtSpot supports a wide variety of SSL types.

Required ports

To use ThoughtSpot webservice securely, ensure that TCP port 443 is open to accommodate incoming connections to ThoughtSpot nodes and clusters.

Configure SSL for web traffic

To add SSL and enable HTTPS in ThoughtSpot, generate the [Certificate Signing Request \(CSR\)](#) [See page 102] and obtain the [SSL certificate chain](#) [See page 104] and the [private key](#) [See page 104].

You can then configure SSL using `tscli` [See page 104] or through the Admin Console [See page 107], and then [test the SSL certificate](#) [See page 108].

Certificate Signing Request

When you generate a CSR, you handle sensitive data. Therefore, ThoughtSpot recommends that its customers generate their own CSRs.

You can generate a CSR in several ways. Most often, you generate a CSR and a new private key [at the same time \[See page 103\]](#). If you already have a private key, [use it to generate a CSR \[See page 103\]](#).

Follow these steps to generate a CSR and a private key. You need a computer you can run Linux commands on, and a recent version of `openssl`.

1. `ssh` into one of your ThoughtSpot nodes.

```
ssh admin@<node_IP>
```

2. Run the command to generate a CSR and private key pair:

```
openssl req -new -newkey rsa:2048 -nodes -out csr.pem  
-keyout pk.key [-subj "/key1=value1/key2=value with space/"]
```

Note the following parameters:

- ThoughtSpot supports a 2048 or 4096 bit key.
- `subj` : a common subject. Logically equivalent to the `-dname` property of `keytool`. Alternatively, you can skip this flag, and `openssl` prompts you to enter this information interactively.
- Optionally, run `add-multivalue-rdn` to allow multiple values to be set for the same key.
- Run `man req` for more details.

If you already have a private key, you can use it to generate a CSR. Follow these steps to generate a CSR with an existing private key:

1. `ssh` into one of your ThoughtSpot nodes.

```
ssh admin@<node_IP>
```

2. Run the command to generate a CSR and private key pair:

```
openssl req -new -key <private_key_file> -nodes -out csr.pem[-subj "/key1=value1/key2=value with space/"]
```

Specify the existing private key file. Refer to the parameters listed above.

SSL certificate chain

The SSL certificate chain must be in PEM format, which is an `X.509v3` file that contains ASCII (Base64) armored data, packed between `BEGIN` and `END` directives. The certificate chain may contain a series of certificates, with the root certificate at the bottom and user-facing, while the ThoughtSpot-specific SSL certificate is at the top.

Private key

The private key must also be in compatible PEM format. It cannot be password-protected, or passphrase-protected.

Note: Do not use a passphrase when creating certificates with ThoughtSpot.

If you are prompted to specify a passphrase, first check if it exists by invoking the following command:

```
openssl rsa -check -in pk.key`
```

If the answer is ‘yes’, you must remove the passphrase first, and then proceed to use the private key with ThoughtSpot.

Configure SSL using tscli

Follow these instructions to install the SSL certificate using tscli:

1. Use the instructions from the certifying authority where you obtained the certificate.

This is usually sent to you by email, or available for download.

2. Copy the certificate and key files to ThoughtSpot:

```
$ scp <key> <certificate> admin@<IP_address>:<certificate-path>
```

3. Log in to the Linux shell using SSH.
4. Change to the directory where you copied the files:

```
$ cd <certificate-path>
```

5. To install the certificate, issue the `tscli` command:

```
$ tscli ssl add-cert <key> <certificate> <IP_address_or_URL>
```

6. To test that the certificate is correctly installed, sign in to the ThoughtSpot application (<https://docs.thoughtspot.com/6.0/admin/setup/logins.html#sign-in-to-the-thoughtspot-application>).

You should see that the application's URL begins with `https://`.

Set the recommended TLS version

ThoughtSpot supports SSL v3, TLS v1.2 by default. Support for TLS v1.0 and v1.1 is included for backwards compatibility. To ensure support for TLS version 1.2:

1. Enable your web browser to support TLS v1.2. This can be done in your browser's advanced settings.
2. Log in to the Linux shell using SSH and run command:

```
tscli ssl tls-status
```

It should respond with

Minimum TLS version supported: 1.2

This will block all usage of older versions.

3. To change this, run `tls ssl set-min-tls-version 1.1` or `tls ssl set-min-tls-version 1.0` as desired for backward compatibility.

Configuration string for load balancers

When enabling SSL support on a load balancer's server-side SSL client profile, make sure to add support for the following ciphers to ensure compatibility between the load balancer and ThoughtSpot.

The following ciphers are currently supported:

```
| TLSv1.2:  
|   ciphers:  
|     TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 - strong  
|     TLS_DHE_RSA_WITH_AES_256_CBC_SHA - strong  
|     TLS_DHE_RSA_WITH_AES_256_CBC_SHA256 - strong  
|     TLS_DHE_RSA_WITH_AES_256_GCM_SHA384 - strong  
|     TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384 - strong  
|     TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 - strong  
|   compressors:  
|     NULL  
|   least strength: strong
```

The cipher string would be as follows:

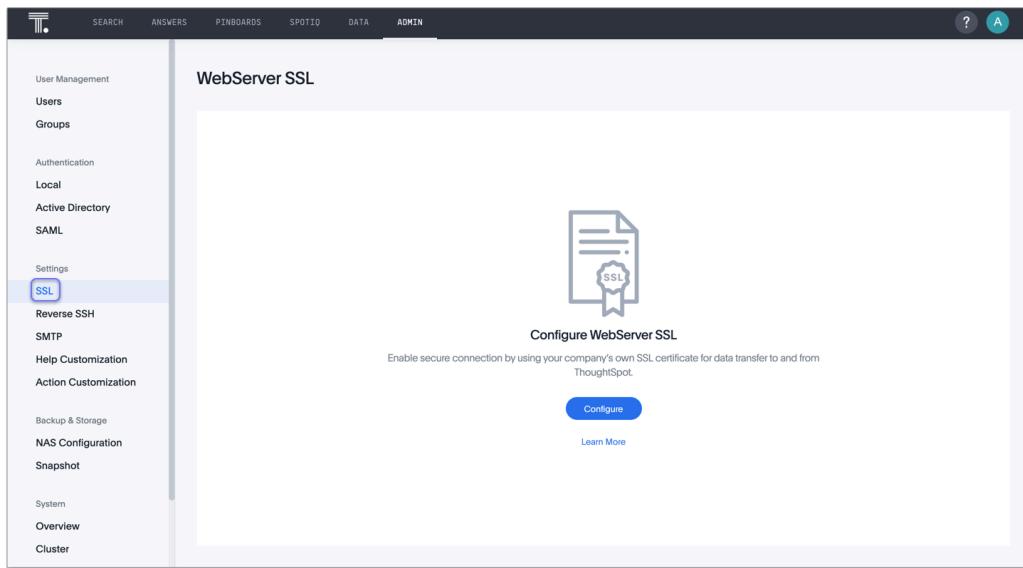
EECDH+AESGCM:EDH+AESGCM:AES256+EECDH: AES256+EDH

You can retrieve these from the ThoughtSpot web server (not against the load balancer) by running the following command on any ThoughtSpot node: `nmap --script ssl-enum-ciphers -p 443 <ThoughtSpot_node_IP_address>`

You must ensure that your load balancer supports these ciphers. If your load balancer cannot support these ciphers, contact ThoughtSpot Support [See page 0].

Configure SSL through the Admin Console

To configure SSL, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SSL** from the side navigation bar that appears.



Click **Configure**.

Fill out the following parameters:

Configure WebServer SSL

SSL Status **1**
 Enable
 Disable

Choose RSA Algorithm * **2**
RSA

Choose Minimum TLS Version * **3**
TLS 1.2

Private Key * **4**
[Browse File](#)

SSL Certificate Bundle * **5**
[Browse File](#)

[Cancel](#) [OK](#)

1 Select **Enable**.

2 Choose **RSA**.

3 Choose **TLS 1.2** as a best practice. ThoughtSpot also supports **TLS 1.1** and **1.0**; set the minimum supported version to **1.1** or **1.0** to use these versions.

4 Attach your private key file.

5 Attach your SSL certificate bundle file.

Click **OK**.

Test the SSL certificate

To test if the certificate is installed correctly, see [Sign in to the ThoughtSpot application \[See page 9\]](#).

Configure SAML

Summary: You can use the Security Assertion Markup Language (SAML) to authenticate users.

ThoughtSpot enables you to use the Security Assertion Markup Language (SAML) to authenticate user. You can set up SAML through the shell on ThoughtSpot [using a `tscli`-based configurator](#) [See page 110], or [through the Admin Console](#) [See page 111]. It is configured to work using service provided by an Identity Provider (IDP).

Configuration prerequisites

Before you configure SAML, collect the following information:

- ThoughtSpot service address [See page 109]
- Service port [See page 109]
- Unique service name [See page 110]
- Skew time in seconds [See page 110]
- IDP Metadata XML File [See page 110]
- Automatically add SAML users to Thoughtspot [See page 110]
- Also use ThoughtSpot internal authentication [See page 110]

ThoughtSpot service address

DNS name of the load balancer *front-end* for multi-node ThoughtSpot clusters, or of the ThoughtSpot server for a single-node ThoughtSpot cluster. If you do not have the DNS name, you can use the front-end IP address. Using the DNS name instead of the IP address is a best practice.

Service port

Service port for ThoughtSpot instance, typically TCP/443.

Unique service name

The unique key ThoughtSpot uses to identify IDP service. Set by the ThoughtSpot Support Team.

The key has the following format: `urn:thoughtspot:callosum:saml`.

Skew time in seconds

Allowed skew time for authentication, or the duration after authentication response is rejected and sent back from the IDP.

Usually set to `3600` seconds.

IDP Metadata XML File

This file is provided by the IDP. The absolute path to the `idp-meta.xml` file is needed for one-time configuration.

Automatically add SAML users to Thoughtspot: (yes/no)

If you choose ‘yes’, then new users will be automatically created in ThoughtSpot upon first successful SSO login.

If you choose ‘no’, then SAML users will not be added in ThoughtSpot upon first successful SSO login. Instead, you must [add users manually \[See page 287\]](#) or through [Active Directory \[See page 118\]](#).

Also use ThoughtSpot internal authentication: (y/n)

If ‘y’, then ThoughtSpot local/internal users (including local administrative users) will still be authenticated outside the scope of SSO.

Configure SAML using tscli

● Note: The configuration persists across updates to newer releases of ThoughtSpot.

To set up SAML on ThoughtSpot for user authentication, follow these steps:

1. Log in to the Linux shell using SSH.
2. Run the `saml configure` command to launch the interactive SAML configuration:

```
tscli saml configure
```

3. Complete the configurator prompts with the information you collected in [Configuration prerequisites \[See page 109\]](#).
4. When the configuration completes, open a browser and navigate to the ThoughtSpot login page. It should show the SSO option.

Configure SAML using the Admin Console

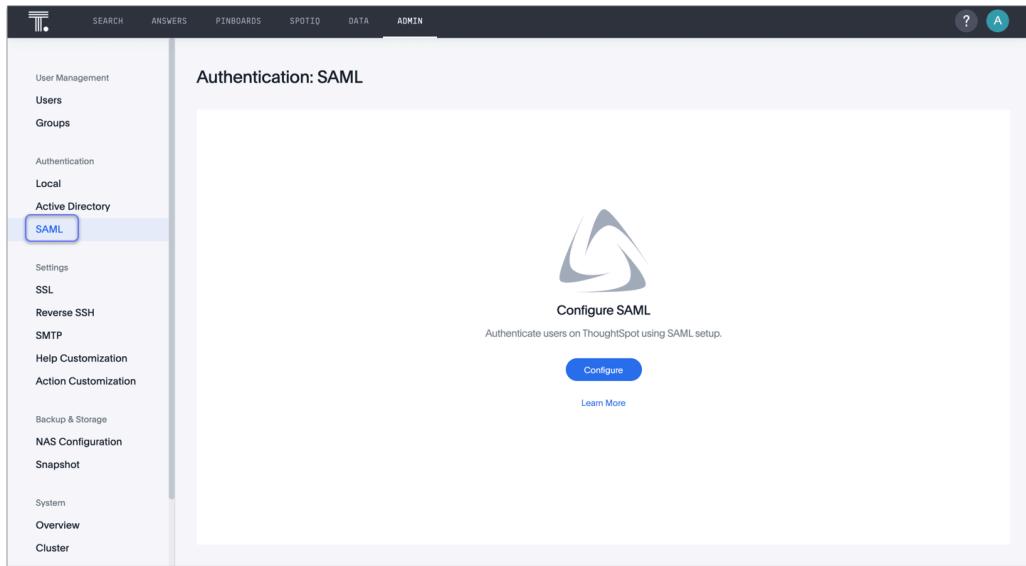
You can use ThoughtSpot's integration with SAML for user authentication. By default, [local authentication \[See page 29\]](#) is enabled. You can also configure a SAML integration with an external Identity Provider (IdP), allowing your ThoughtSpot users to log in using one of the supported Identity Providers: Okta, Ping Identity, CA SiteMinder, or ADFS.

You can configure the SAML integration through the Admin Console.

Note: If you configure authentication through SAML, you cannot also configure authentication through Active Directory.

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **SAML** from the side navigation bar that appears.

Configure SAML



Click the **Configure** button in the middle of the screen.

Configure SAML

ThoughtSpot Service Address *
Supply fully qualified and resolvable domain name for ThoughtSpot service

Port *

Unique Service Name *
Unique service id will be used as IDP to identify client

Skew Time in Seconds *
Suggested skew time is 3600 seconds

Protocol *

IDP Metadata XML File *
[Browse File](#)

Automatically add SAML users to ThoughtSpot upon first authentication

Yes
 No

[Cancel](#) [OK](#)

Fill in the following parameters:

1. **ThoughtSpot Service Address:** A fully qualified and resolvable domain name for the ThoughtSpot service. For example, *thoughtspot.thoughtspot-customer.com*.
2. **Port:** Port of the server where your ThoughtSpot instance is running. For example, port `443`.
3. **Unique Service Name:** The unique key used by your Identity Provider to identify the client. For example, *urn:thoughtspot:callosum:saml*.

4. **Skew Time in Seconds:** The allowed skew time, after which the authentication response is rejected and sent back from the IDP. 86400 is a popular choice. The default is 3600.
5. **Protocol:** The authentication mechanism for ThoughtSpot. For example, `http` or `https`.
6. **IDP Metadata XML File:** The absolute path to your Identity Provider's metadata file. This file is provided by your IDP. You need this file so that the configuration persists over upgrades. It is a best practice to set it up on persistent/HA storage (NAS volumes) or in the same absolute path on all nodes in the cluster. For example, `idp-meta.xml`.
7. **Automatically add SAML users to ThoughtSpot upon first authentication:** Choose whether or not to add SAML users to ThoughtSpot when they first authenticate. If you choose 'yes', then new users will be automatically created in ThoughtSpot upon first successful SSO login. If you choose 'no', then SAML users will not be added in ThoughtSpot upon first successful SSO login. Instead, you must [add users manually](#) [See page 25].

After you fill in all parameters, click **OK**.

❶ Note: ThoughtSpot adds external users, or users that authenticate through SAML or Active Directory, to the **all** group by default. This group has no privileges. You must manually assign users to ThoughtSpot groups to give them privileges, such as **can upload user data**, or **can manage data**.

Enable SSH access through Active Directory

Summary: You can enable Active Directory (AD) access on each ThoughtSpot node.

Enable user access on a ThoughtSpot node through AD

In ThoughtSpot, you can enable Active Directory (AD) access on specified nodes of the cluster. To enable AD access on the entire cluster, enable AD access on each node. You must similarly configure all new nodes that you subsequently add to the cluster.

To enable system AD user access, issue the `tscli sssd enable` [See page 0] command:

```
tscli sssd enable --user <USER> --domain <DOMAIN>
```

Here, `USER` is the user account you are enabling on the domain `DOMAIN`.

The system prompts for password credentials before completing the step.

Note: The user must have permission to join a computer or VM to the domain.

Set sudo AD Group on a ThoughtSpot node

Just like enabling AD-based access on a node, you can enable `sudo` AD groups only on specific nodes.

To set AD group `sudo` access for the entire cluster, you must run the command on each node. You must similarly configure all new nodes that you subsequently add to the cluster.

To enable `sudo` permissions for AD group, issue the `tscli sssd set-sudo-group` [See page 0] command:

```
tscli sssd set-sudo-group <ACTIVE_DIRECTORY_GROUP_NAME>
```

Clear sudo AD Group on a ThoughtSpot node

You may clear the `sudo` AD group only on the node where you run the command, not for the entire cluster. To clear access for the entire cluster, you must run the command on each node. You must similarly configure all new nodes that you subsequently add to the cluster.

To clear `sudo` permissions for the AD group, issue the [tscli sssd clear-sudo-group \[See page 0\]](#) command:

```
tscli sssd clear-sudo-group <ACTIVE_DIRECTORY_GROUP_NAME>
```

Disable AD access on a ThoughtSpot node

You can disable AD access, individually on each node of the cluster.

To disable system AD user access, issue the [tscli sssd disable \[See page 0\]](#) command:

```
tscli sssd disable
```

Note: Running this command also removes the AD group from sudo list.

Related Information

- The [sssd command \[See page 0\]](#) in the `tscli` command reference

About LDAP integration

Summary: You authenticate users against an LDAP server.

Some companies use LDAP (Lightweight Directory Access Protocol) to manage user authentication.

Using LDAP provides security and makes user management more centralized. You can choose to authenticate users against an LDAP server, against ThoughtSpot, or both.

ThoughtSpot supports both anonymous and non-anonymous LDAP integration. Non-anonymous LDAP binding is more rigorous than anonymous authentication, but it should help you track what your users are querying and keep a log trace for auditing purposes.

If you have been using ThoughtSpot with users you created manually, and you now want to transition to LDAP, please contact ThoughtSpot Support. They can assist you in migrating existing users to their LDAP equivalents.

ThoughtSpot supports LDAP with [Active Directory \[See page 118\]](#).

On ThoughtSpot version 6.3 and later, and on version 6.0.5, ThoughtSpot also integrates with OpenLDAP for user authentication. LDAP provides security and makes user management more centralized. To enable user authentication through OpenLDAP, [contact ThoughtSpot Support \[See page 0\]](#).

Configure authentication through Active Directory

Summary: ThoughtSpot enables you to set up integration with LDAP using Active Directory. After successful setup, you can authenticate users against a secure LDAP server.

Note: This article describes how to set up integration with LDAP using Active Directory. On ThoughtSpot version 6.3 and later, and on release 6.0.5, ThoughtSpot also integrates with OpenLDAP for user authentication. LDAP provides security and makes user management more centralized. To enable user authentication through OpenLDAP, contact ThoughtSpot Support.

Configuration prerequisites

You can configure LDAP [using tscli](#) [See page 120] or [through the Admin Console](#) [See page 121].

Before you configure ThoughtSpot for Active Directory, collect the following information:

- URL [See page 118]
- Domain name [See page 119]
- Search base [See page 119]
- SSL [See page 119]
- Automatically add LDAP or AD users in ThoughtSpot? [See page 119]
- Also use ThoughtSpot internal authentication? [See page 120]

URL

Required to connect to Active Directory.

For example, `ldap://ad.yourdomain.local:389` or `ldap://ad.yourdomain.local:636`

Domain name

Default domain under which users who want to be authenticated against Active Directory reside. When a user logs in with a username, the default domain is added to the username before sending it to the LDAP server. If users reside in multiple sub-domains, you can still designate one of them as the default.

Authentication against multiple domains is not supported.

Users who don't belong to the default domain must explicitly qualify their username when they log in.

For example: `username@ad.yourdomain.local`

Search base

LDAP search base. The scope of searching user information, like *email* and *Display name*, within AD.

SSL

If you want to use SSL, you must obtain the SSL certificate from an issuing authority.

If AD servers are behind a load balancer, you must procure the SSL certificate to identify ThoughtSpot to the load balancer. The communication after the load balancer is non-secure. ThoughtSpot does not support a scenario where multiple AD servers provide their own SSL certificates.

Automatically add LDAP or AD users in ThoughtSpot? (yes/no)

If you choose 'yes', new users are automatically created within ThoughtSpot when successfully authenticated against AD. ThoughtSpot doesn't cache passwords for AD-authenticated users.

If you choose 'no', users have to be manually created with a dummy password as a placeholder in ThoughtSpot before they can log in. The username you specify when creating the LDAP-authenticated user manually in ThoughtSpot has to be domain qualified, for example:

`username1@ad.yourdomain.local`.

In order to log in to ThoughtSpot, the user has to exist in ThoughtSpot independent of whether that user is authenticated against AD or against ThoughtSpot's internal authentication.

Also use ThoughtSpot internal authentication? (yes/no)

If you choose ‘yes’, ThoughtSpot will first attempt to authenticate the user against AD. If that attempt fails, it will then attempt to authenticate the user as an internal/local ThoughtSpot user. If either of these succeed, then the user is successfully logged in. This is useful in scenarios where some users are not in AD and are created only in ThoughtSpot.

Configure LDAP using tscli

You do not have to create a user called `tsadmin` on your LDAP server. Internal authentication can be used for `tsadmin`. To configure AD based authentication, follow these steps:

1. Log in to the Linux shell using SSH.
2. Run the command to configure AD authentication:

```
$ tscli ldap configure
```

3. Answer the prompts using the information you collected under **Before you begin** section. For example:

```
Choose the LDAP protocol:  
[1] Active Directory  
Option number: 1  
Configuring Active Directory  
URL to connect to Active Directory. (Example: ldap://ad.yourdomain.local:389): ldaps://ad.yourdomain.local:636  
Default domain (Example: ldap.thoughtspot.com): yourdomain.local  
Use SSL (LDAPS) (y/n): n  
LDAP search base (Example: cn=Users): cn=Users,ou=organit,dc=youdomain,dc=local  
Automatically add LDAP users in ThoughtSpot (y/n): y  
Also use ThoughtSpot internal authentication (y/n): y
```

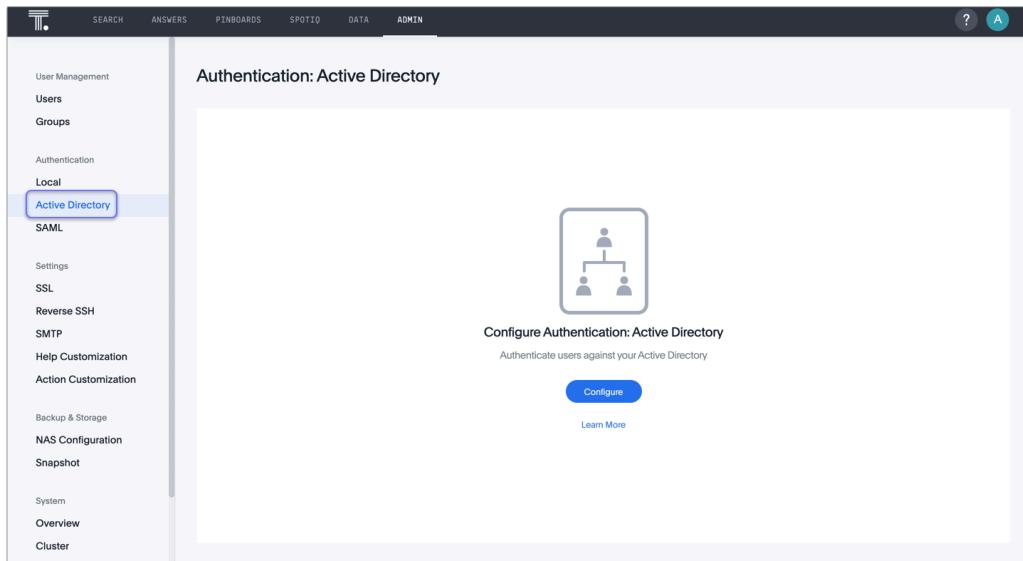
4. If you are using SSL, [add the SSL certificate for AD \[See page 125\]](#).

5. If you want to remove the AD configuration, issue the following command:

```
$ tscli ldap purge-configuration
```

Configure LDAP using the Admin Console

Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Active Directory** from the side navigation bar that appears.



Click the **Configure** button in the middle of the screen, and add your information based on the information you collected in the [prerequisites \[See page 118\]](#).

Note: If you configure authentication through Active Directory, you cannot also configure authentication through SAML.

Configure Authentication: Active Directory

URL *

Domain Name Preferred *

Search Base *

Automatically add Active directory users to ThoughtSpot upon first successful authentication

Yes
 No

SSL Required

Yes
 No

Cancel OK

If you select **yes** for **SSL required**, ThoughtSpot prompts you to enter the SSL [certificate](#) [See page 119] and certificate alias. The certificate alias is a unique name you make up for your SSL certificate, in string format. If you replace the SSL certificate, you need a new certificate alias.

Configure Authentication: Active Directory

URL *

Domain Name Preferred *

Search Base *

Automatically add Active directory users to ThoughtSpot upon first successful authentication

Yes
 No

SSL Required

Yes
 No

SSL Certificate *

Could not pull SSL Certificate from AD server, upload certificate.

[Browse File](#)

Certificate Alias *

[Cancel](#) [OK](#)

After you add all your information, click **OK**.

Note: ThoughtSpot adds external users, or users that authenticate through SAML or Active Directory, to the **all** group by default. This group has no privileges. You must manually assign users to ThoughtSpot groups to give them privileges, such as **can upload user data**, or **can**

manage data.

Add the SSL certificate for LDAP

Summary: Install the certificate to support LDAPS.

When you set up LDAP, you specified whether or not to use SSL for LDAP (LDAPS). If using SSL, you must install the LDAP SSL certificate. Before you can add the SSL certificate, you must [Configure LDAP for Active Directory \[See page 118\]](#).

You must have the SSL certificate before you start. For more information on obtaining an SSL certificate, see [Configure SSL \(secure socket layers\) \[See page 102\]](#).

To add the SSL certificate for LDAP:

1. Follow the instructions from your certifying authority to obtain the certificate. This is usually sent by email or available for download.
2. Copy the certificate to ThoughtSpot:

```
$ scp <certificate> admin@<IP_address>:<path>
```

3. Log in to the Linux shell using SSH.
4. Change directories to where you copied the certificate:

```
$ cd <path>
```

5. Run the command to configure SSL for LDAP, designating an alias for this certificate using the `<name>` parameter:

```
$ tscli ldap add-cert <name> <certificate>
```

Test the LDAP configuration

Summary: This procedure allows you to test the LDAP connection you created.

After configuring LDAP, you can test to make sure it is working by issuing a command.

1. Log in to the Linux shell using SSH.
2. Issue the LDAP testing command, supplying the information for the LDAP server you configured, as in this example:

```
$ ldapsearch -x -h 192.168.2.61 -p 389 -D "testuser@laptop.thoughtspot.com" -W -b "dc=ldap,dc=thoughtspot,dc=company" cn
```

3. Supply the LDAP password when prompted.
4. If the connection works, you can see a confirmation message.

Sync users and groups from LDAP

Summary: Use this procedure to synchronize your ThoughtSpot system with an LDAP server through Active Directory.

Prerequisites

Before synchronizing users and groups, [set up integration with LDAP using Active Directory \[See page 118\]](#). Then, collect the following information:

- IP address and port of the server where your ThoughtSpot instance is running.

This hostport must be in the following format `http(s)://<host>:<port>` or
`http(s)://<domain>`.

- Administrator login username and password for your ThoughtSpot instance.
- URL of the LDAP server, or hostport.

For example, `ldap://192.168.2.48:389`

- Login username and password for the LDAP system.

An example username would be `moo_100@ldap.thoughtspot.com`

- Distinguished Name (DN) for the base to start searching for users in the LDAP system.

For example, `DC=ldap,DC=thoughtspot,DC=com`

- The Python synchronization script, in case you want to modify it or create your own, is here:

```
/usr/local/scaligent/release/callosum/utilities/ldap_sync_python_api/syncUsersAndGroups.py
```

Fetch users and groups from LDAP with Active Directory

There are two ways for you to fetch users and groups from LDAP and populate them into your ThoughtSpot system:

- Run the synchronization script in interactive mode, which walks you through the process (shown here).
- Create your own Python script by using the ThoughtSpot Python APIs. If you need details on the Python APIs, [contact ThoughtSpot Support \[See page 0\]](#). If you choose this method, you can run the script periodically using a cron job.

Note: When you run the synchronization script, you perform a one-time sync. You must schedule a recurring sync using a cron job or your own scheduling tool to keep your ThoughtSpot users up to date with your users in LDAP.

Run the sync script

To run the LDAP sync script in interactive mode:

1. Log in to the Linux shell using SSH.
2. Run the command to start the script:

```
python3 syncUsersAndGroups.py interactive
```

3. Answer the prompts using the information you collected above. Specify 2 for the scope if you would like to sync all groups, including subgroups. For example:

```
Complete URL of TS server in format "http(s)://<host>:<port>": http://10.77.145.24:8088
Disable SSL authentication to TS server (y/n): y
Login username for ThoughtSpot system: admin
Login password for ThoughtSpot system: 12345
Complete URL of server where LDAP server is running in
format ldap(s)://<host>:<port>: ldap://192.168.2.48:389
Login username for LDAP system: moo_100@ldap.thoughtspo
t.com
Login password for LDAP system: 12345
Syncs user and groups between LDAP and TS systems (y/
n): y
Delete entries in ThoughtSpot system that are not curre
ntly in LDAP tree being synced (y/n): n
Distinguished name for the base to start searching grou
ps in LDAP System: DC=ldap,DC=thoughtspot,DC=com
Scope to limit the search to (choice number)
0:base Searching only the entry at the base DN
1:one Searching all entries on level under the base DN
- but not including the base DN
2:tree Searching of all entries at all levels under an
d including the specified base DN: 2
```

Filter string to apply the search to: (|(CN=TestGroupAl
pha)(CN=TestGroupBeta))

Answering this prompt is optional. If left blank, the default value of '(CN=*)' will be used.

Apply sync recursively, i.e. Iterates through group mem
bers and creates member groups, users and relationships
in a recursive way. (y/n): n

This prompt is asking if you would like to include group members even if they do not belong to the current sub tree that is being synced.

4. Alternatively, to input your own shorthand script commands:

Issue the Python script commands, supplying all this information, following this format
example:

```
python3 syncUsersAndGroups.py script \
--ts_hostport <ts_hostport> \
--disable_ssl \
--ts_uname <ts_username> \
--ts_pass <ts_password> \
--ldap_hostport '<ldap_hostport>' \
--ldap_uname '<ldap_username>' \
--ldap_pass '<ldap_password>' \
--sync \
--purge \
--basedn 'DC=ldap,DC=thoughtspot,DC=com' \
--filter_str '(|(CN=TestGroupAlpha)(CN=TestGroupBeta))' \
--include_nontree_members \
--user_identifier <user_identifier> \
--authdomain_identifier <authdomain_identifier> \
--email_identifier <email_identifier>
```

The bottom half of the preceding command targets sub trees under the DC called TestGroupAlpha and TestGroupBeta, and iterates through them recursively to create/sync users, groups, and their relationships in the ThoughtSpot system. It also deletes any other entities created in the ThoughtSpot system from this LDAP system that are not currently being synced.

syncUsersAndGroups.py command-line switches

The following table provides a description of each command-line switch available for the `syncUsersAndGroups` python script.

Switch	Description
<code>--ts_hostport <ts_hostport></code>	ThoughtSpot cluster host port. Default port is 8088.
<code>--disable_ssl</code>	Controls the communication between the sync script and the ThoughtSpot cluster. It disables SSL communications between the script and the cluster ONLY, and prevents the need to provide SSL certs during the script execution in order to create users and groups.
<code>--ts_uname</code>	ThoughtSpot cluster username. The <code>admin</code> user is usually used.
<code>--ts_pass</code>	ThoughtSpot cluster password.

--ldap_hostport	AD/LDAP server port that is queried. Default is 389.
--ldap_uname	Username for the LDAP/AD server.
--ldap_pass <ldap_pass-word>	Password for the LDAP/AD server.
--sync	Syncs users and groups which match the <code>basedn</code> and <code>filter_str</code> queries to your ThoughtSpot cluster.
--purge	Purges any users that exist in ThoughtSpot, but not in AD.
--basedn	Place in the directory that will be searched for users.
--filter_str	Further filters results from your base DN.
--include_nontree_members	Includes group members from LDAP/AD even if they do not belong to the current subtree that is being synced.
--user_identifier <user_identifier>	User name identifier key for user creation or sync.
--authdomain_identifier <authdomain_identifier>	Override domain name to be appended to user identifier in user name.
--email_identifier <email_identifier>	Email identifier key for user creation or sync.

Configure NAS file system

Summary: Some operations, like backup, restore, and data loading, require either reading or writing very large files. You can mount a network attached storage (NAS) file system to support these operations. Your NAS storage can be in the drive format you choose.

About NAS mount

ThoughtSpot enables you to mount a NAS file system for storing or accessing large files. The file system mounts at the same location on each node in the cluster automatically. When any node restarts, the file system mounts again automatically, if it can be found.

When supplying a directory for writing or reading a backup, you can specify a new mount point within `/export` as the directory to use. Likewise, you can stage data there for loading. It is best to have 2 separate NAS volumes, individually dedicated to data loads and backups.

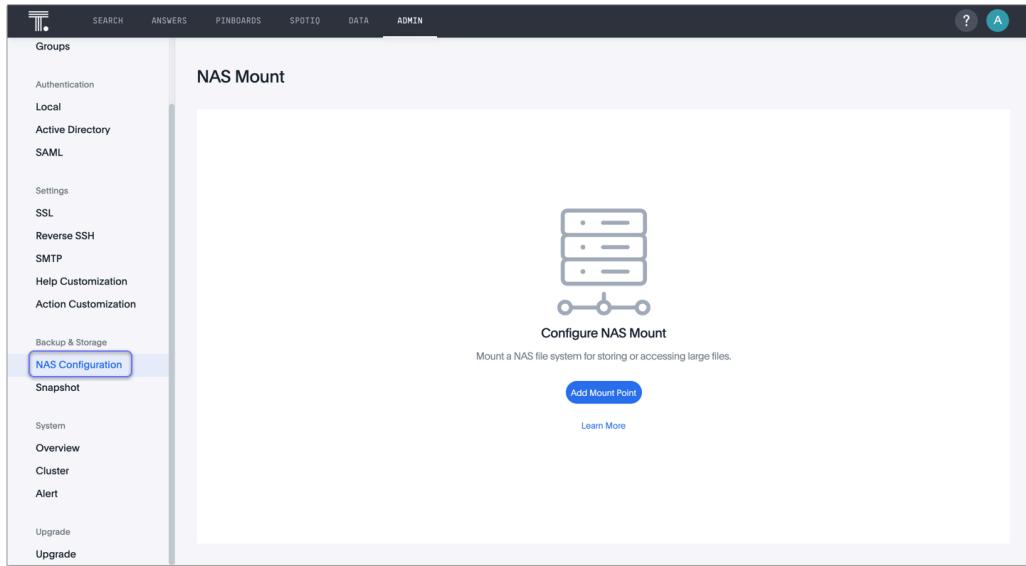
Backups are written by the Linux user `admin`. If that user does not have permission to write to the NAS file system, you can write the backups to a disk (for example `/export/sdc1`, `/export/sdd1`, `/export/sde1`, or `/export/sdf1`) and then set up a cron job that executes as root user and copies the backup to the NAS device every night, then deletes it from the directory.

Do not send the periodic backups or stage files on `/export/sdb1` since it is a name node. It is used internally by Hadoop Distributed File System (HDFS) and if this drive fills up, it can cause serious problems. Do not allow backups or data files to accumulate on ThoughtSpot. If disk space becomes limited, the system will not function normally.

You can mount NAS [through the Admin Console \[See page 132\]](#), or [using tscli \[See page 135\]](#).

Mount NAS using the Admin Console

To mount a NAS file system using the Admin Console, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **NAS Configuration** from the side navigation bar that appears.



1. Select **Add Mount Point**.
2. Specify the parameters in the dialog box that appears:

Add Mount Point

Mount Type *

NFS 1

CIFS

Server Address * 2

Path on Server * 3

Local Mount Point * 4

Username * 5

Password * 6

Optional Mount Parameters

Other command-line options are available to forward to the command. 7

- 1 Choose the mount type: either NFS (Network File System) or CIFS (Common Internet File System).
- 2 Specify the server address, in the form `storageservername.file.yourdomain.net`.
- 3 Specify the server path.
- 4 Specify the local mount point. This should be within the `/export` directory.
- 5 If you choose the CIFS mount type, you must specify a username and password.
- 6 If you choose the CIFS mount type, you must specify a username and password.

- 7 Optionally specify additional command-line subcommands and flags. Refer to the tscli reference [See page 0] for NAS commands.

3. Click **OK**.

You can unmount NAS from the Admin Console, or by running `tscli nas unmount --dir <directory>`.

Mount NAS using tscli

To mount a NAS file system using the tscli, follow these steps:

1. Log in to the Linux shell using SSH.
2. Mount the directory to the file system by issuing the appropriate command:
 - Example for an NFS (Network File System) directory:

```
tscli nas mount-nfs --server storageservername.firebaseio.yourdomain.net  
--path_on_server <path> /tsdev-backup --mount_point /export/BACKUPS/  
--options vers=<version>,sec=<security scheme>,<OPTIONS>
```

Note: Other command-line options are available to forward to the command (default: noexec).

- Example for a CIFS (Common Internet File System) directory. Use `1001` for the `uid` and `gid`, as in the example:

```
tscli nas mount-cifs --server storageservername.firebaseio.yourdomain.net  
--path_on_server /tsdev-backup --mount_point /export/BACKUPS/  
--username 'avtprdweutspotdev' --uid 1001 --gid 1001 --options 'vers=3.0'
```

Note: Other command-line options are available to forward to the `mount.cifs` command (default: `noexec`).

3. Use the mounted file system by referring to its mount point.
4. When you are finished with it, you can optionally unmount the NAS file system:

```
tscli nas unmount --dir <directory>
```

Set up monitoring

Summary: Setting up monitoring is a one time operation.

To configure monitoring of your cluster, set up the frequency of heartbeat and monitoring reports and an email address to receive them.

1. Log in to the Linux shell using SSH.
2. Issue the `tscli` command to set up monitoring:

```
tscli monitoring set-config  
  --email <email>  
  --heartbeat_interval <heartbeat_interval>  
  --report_interval <report_interval>
```

The parameters are:

- `--email <email>` is a comma separated list (no spaces) of email addresses where the cluster will send monitoring information.
- `--heartbeat_interval <heartbeat_interval>` is the heartbeat email generation interval in seconds. Must be greater than 0.
- `--report_interval <report_interval>` sets the cluster report email generation interval in seconds. Must be greater than 0.

3. To view your settings and verify that they have been applied, issue:

```
tscli monitoring show-config
```

You should see information like:

Monitoring Configuration:

Alert Email: dev-alerts@thoughtspot.com

Heartbeat Interval: 900 sec

Report Interval: 21600 sec

4. After the heartbeat interval has passed, check your email to verify that emails are being delivered.
5. If you don't receive any emails, [verify that email is working \[See page 0\]](#).

Configure support services

Summary: There are several configurations you can set up in your installation to ensure your company's support from ThoughtSpot works smoothly.

Set up a reverse tunnel for support

You can set up a reverse tunnel to allow ThoughtSpot Support to get access to your ThoughtSpot instance, to perform support-related activities. This setup is scalable, more secure, and a much simpler alternative to using a virtual meeting room.

Granting remote support access can streamline troubleshooting activities, because it enables your support agent to work directly on your computer from a secure setting. The remote tunnel enables SSH and HTTP access to your ThoughtSpot instance by ThoughtSpot Support. This access can be granted and revoked easily, so you can enable it for a troubleshooting session, and then disable it again. Before doing this procedure, make sure your company's security policies allow reverse tunneling.

You can set up a reverse tunnel [using tscli \[See page 139\]](#), or through the Admin Console [\[See page 141\]](#).

Note: Before you set up a reverse tunnel, open port 22 in your firewall outgoing rules to add `tunnelrelay.thoughtspot.com` to your list of allowed domains.

Using remote support with tscli

To enable remote support, follow these steps:

1. Contact [ThoughtSpot Support \[See page 0\]](#) and open a support ticket for making the appropriate reverse tunnel settings on our end. Provide the cluster name of the cluster for which you want to enable remote support.
2. After the ticket is completed, continue with the remaining steps in this procedure to make the settings on your side.
3. Log in to the Linux shell using SSH.
4. Issue the command to configure the destination for the remote tunnel.

You only need to do this one time, when you are enabling the tunnel for the very first time.

After that, this setting persists when you start and stop the remote tunnel.

```
$ tscli support set-remote --addr tunnelrelay.thoughtspot.com --user ubuntu
```

5. Test that the setting is configured:

```
$ tscli support show-remote
```

6. Enable the remote tunnel:

```
$ tscli support start-remote
```

7. Contact ThoughtSpot Support (<https://docs.thoughtspot.com/latest/admin/misc/contact.html>) and test your setup.
8. After your remote session with ThoughtSpot Support is over, turn the remote tunnel off until you need to use it again:

```
$ tscli support stop-remote
```

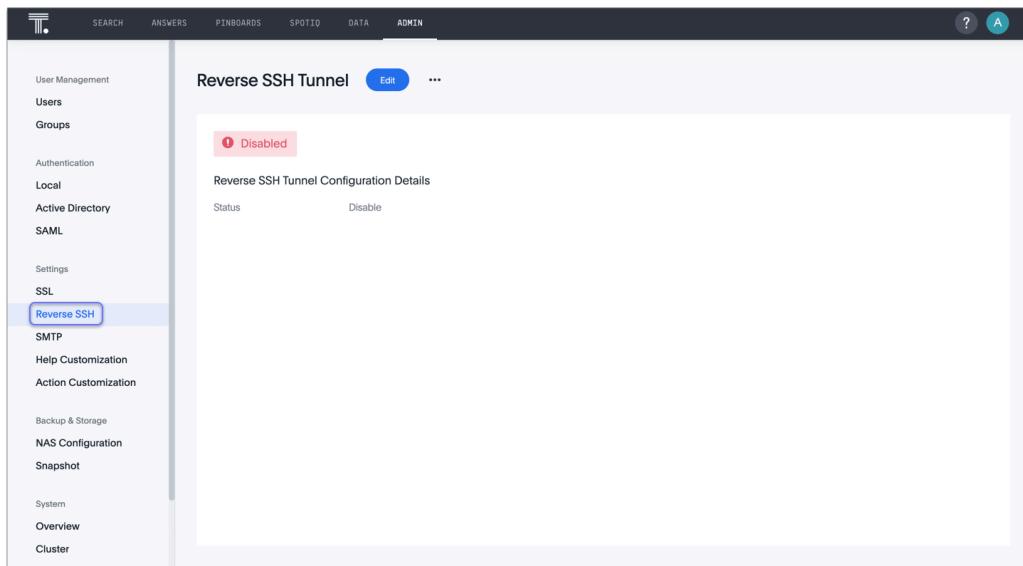
You can repeat the steps to start and stop the remote tunnel as needed for future support operations.

9. Ensure that the remote tunnel is disabled:

```
$ tscli support show-remote
```

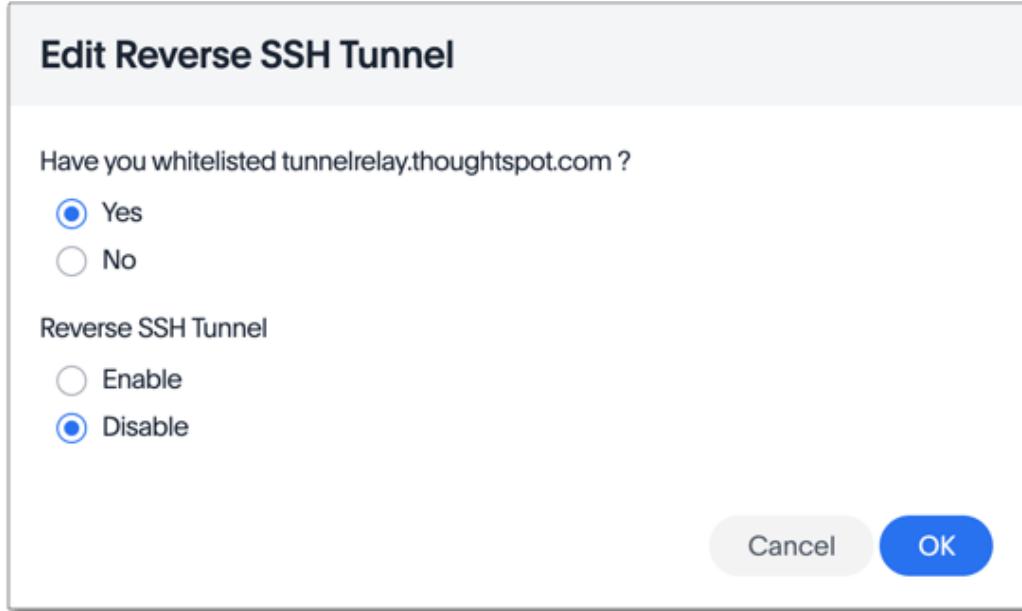
Configure reverse SSH tunnel with the Admin Console

To configure a reverse SSH tunnel, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Reverse SSH** from the side navigation bar that appears.



Select **Edit** to enable or disable the tunnel.

The **Edit Reverse SSH Tunnel** dialog box appears.



The system asks if you have added `tunnelrelay.thoughtspot.com` to your list of allowed domains.

If you select **no**, you cannot enable or disable the reverse SSH tunnel:



If you have added `tunnelrelay.thoughtspot.com` to your list of allowed domains, by opening port `22`, the system allows you to enable or disable the reverse SSH tunnel.



Make the necessary changes and select **OK**.

Configure a secure file server

ThoughtSpot Support uses a secure file server to distribute new releases and to access your logs and troubleshooting files. You must also use a secure server connection to enable the optional performance statistics collection.

Before uploading files to the secure file server, obtain your user name and password for logging into the secure file server. You can get these from ThoughtSpot Support.

Configuring the connection to the file server is a one-time operation. You do not have to reconfigure the connection unless your password changes. Note that you can do a one time override of the user and password you used to configure the connection. This is done by passing a different user and password on the command line when uploading or downloading a file.

To configure the connection to the secure file server, follow these steps:

1. Log in to the Linux shell using SSH.
2. Issue the command to configure the file server:

```
$ tscli fileserver configure --user <user_name> [--password <password>]
```

If you do not supply the `--password` parameter, you are prompted to enter it.

Call home with cluster usage data

“Call home” data is metadata and usage data from your ThoughtSpot cluster. This data allows ThoughtSpot’s Support team to troubleshoot your cluster. They use the data to see basic usage information over time for your ThoughtSpot instance. ThoughtSpot’s “call home” functionality intermittently sends a call home bundle of statistics to a ThoughtSpot S3 server through HTTPS. The data is encrypted at rest on the server.

By default, call home is enabled on your cluster. You can disable call home in by doing the following:

1. Log in to the ThoughtSpot server as `admin` user.
2. Use the `tscli` command to disable.

```
$ tscli callhome disable
```

Designate a support contact

A support contact person can answer your questions about data and search at your company. If the person can’t answer your system and software-related questions, that person should submit the questions to ThoughtSpot Support. The designated support contact should have an available email and phone number.

To designate the custom support contact, follow these steps:

1. Log in to the Linux shell using SSH.
2. Issue the `tscli` command to set the email address:

```
$ tscli support set-admin-email <email_address>
```

3. Issue the `tscli` command to set the phone number:

```
$ tscli support set-admin-phone <phone_number>
```

4. If you need to reset both of these to the default (ThoughtSpot Support), issue these commands:

```
$ tscli support rm-admin-email  
$ tscli support rm-admin-phone
```

Manage the feedback contact

Users in ThoughtSpot may be asked for feedback for new or Beta features in the system. By default, feedback goes directly to ThoughtSpot support.

Alternatively, and especially in cases of using ThoughtSpot in embedded mode, you can send feedback to someone in your company.

Your designated feedback contact should have an available email. To designate the custom feedback contact, follow these steps:

1. Log in to the Linux shell using SSH.
2. To set the feedback email address, issue this command:

```
$ tscli support set-feedback-email <email_address>
```

3. Verify the email address is set:

```
$ tscli support show-feedback-email
```

To reset the email to the default (ThoughtSpot support), issue this command:

```
$ tscli support rm-feedback-email
```

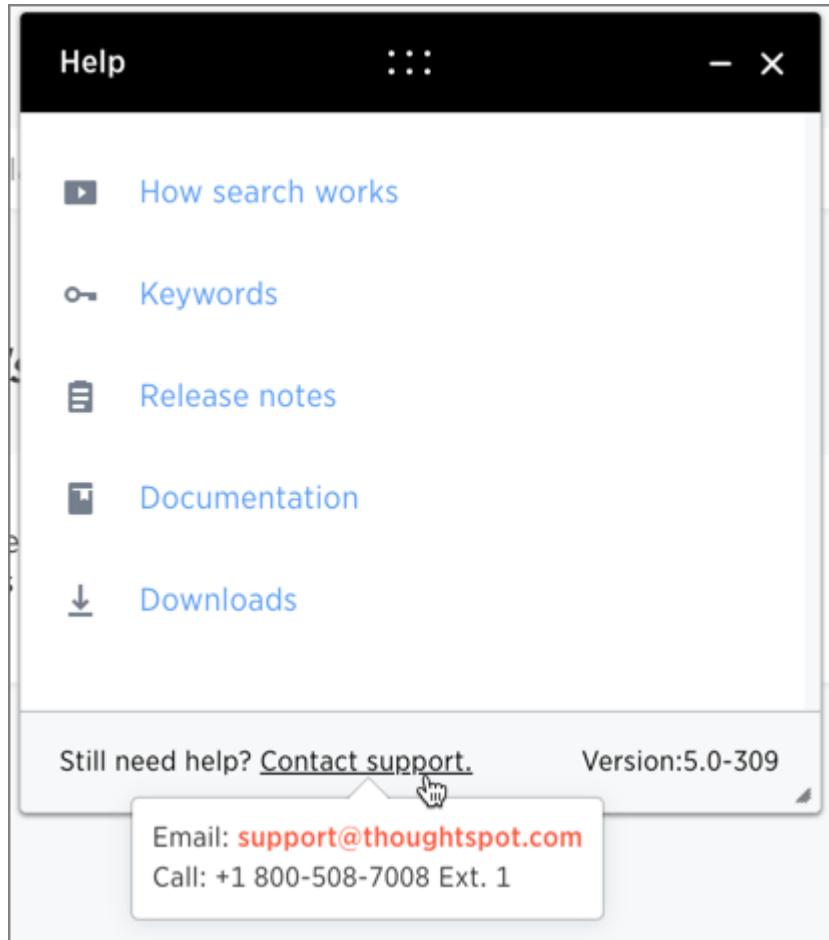
You can also choose not to send feedback on your system. Issue this command:

```
$ tscli support set-feedback-email ''
```

How users find your company's support contact

After you set the custom support contact information, your users can see it in the following parts of ThoughtSpot:

- In the Help Center, when a user selects **Contact Support**.



- In error messages, when a user selects **What Happened?**

Network ports

Summary: Note the list of ports, both required and optional, for regular operations of ThoughtSpot.

For regular operations and for debugging, there are some ports you must keep open to network traffic from end users. Another, larger list of ports must be kept open for network traffic between the nodes in the cluster.

Ports for operations and debugging

The following ports must be open for requests from users.

Port	Protocol	Service Name	Direction	Source	Destination	Description
22	SSH	SSH	bidirectional	Administrators IP addresses	All nodes	Secure shell access. Also used for scp (secure copy).
443	HTTPS	HTTPS	bidirectional	All users IP addresses	All nodes	Secure HTTP.
12345	TCP	Simba	bidirectional	Administrators IP addresses	All nodes	Port used by ODBC and JDBC drivers when connecting to ThoughtSpot.

Ports for intracluster network operations

Static ports are used for communication between services within the cluster. ThoughtSpot recommends that you open all ports within a cluster. This is not required, but it will ensure that cluster communication works properly if additional ports are used in a future software release.

If your organization does not allow you to open all ports, make sure you open the required intracluster ports listed in the following table. In addition, a number of ports are dynamically assigned to services, which change between runs. The dynamic ports come from the range of ports that are dynamically allocated by Linux (20K+).

Port	Protocol	Service Name	Direction	Source	Dest.	Description
80	TCP	nginx	inbound	All nodes	All nodes	Primary app HTTP port (nginx)
443	TCP	Secure nginx	inbound	All nodes	All nodes	Primary app HTTPS port (nginx)
2100	RPC	Oreo RPC port	bidirectional	All nodes	All nodes	Node daemon RPC
2101	HTTP	Oreo HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Node daemon HTTP
2181	TCP	Zookeeper servers listen on this RPC port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this RPC port for client connections
3181	TCP	Zookeeper servers listen on this RPC port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this RPC port for client connections
4181	TCP	Zookeeper servers listen on this RPC port for client connections	bidirectional	All nodes	All nodes	Zookeeper servers listen on this RPC port for client connections
2200	RPC	Orion master RPC port	bidirectional	All nodes	All nodes	Internal communication with the cluster manager
2201	HTTP	Orion master HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the cluster manager
2205	TCP	Cluster update service TCP port	bidirectional	All nodes	All nodes	Internal communication with the cluster manager

Port	Protocol	Service Name	Direction	Source	Dest.	Description
2210	RPC	Cluster stats service RPC port	bidirectional	All nodes	All nodes	Internal communication with the stats collector
2211	HTTP	Cluster stats service HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the stats collector
2230	RPC	Callosum stats collector RPC port	bidirectional	All nodes	All nodes	Internal communication with the BI stats collector
2231	HTTP	Callosum stats collector HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the BI stats collector
2240	RPC	Alert manager	bidirectional	All nodes	All nodes	Port where alerting service receives alert events
2241	RPC	Alert manager	bidirectional	All nodes	All nodes	Port where alerting service receives alert events
2888	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
3181	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
3888	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4000	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4001	HTTP	Falcon worker HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the data cache
4002	HTTP	Falcon worker HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the data cache
4003	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4004	RPC	Falcon worker RPC port	bidirectional	All nodes	All nodes	Port used by data cache for communication between themselves
4021	RPC	Sage metadata service port (exported by Tomcat), Callosum services like meta-data services, meta-data-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts metadata service for metadata
4181	RPC	Ports used by Zookeeper servers for communication between themselves	bidirectional	All nodes	All nodes	Ports used by Zookeeper servers for communication between themselves
4201	HTTP	Sage auto complete server HTTP interface port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4231	HTTP	Sage index server HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service
4232	RPC	Sage index server metadata subscriber port	bidirectional	All nodes	All nodes	Port used for search service internal com- munication
4233	RPC	Sage index server RPC port	bidirectional	All nodes	All nodes	Port used for search service internal com- munication
4241	HTTP	Sage auto complete server HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Port used to debug the search service
4242	RPC	Sage auto complete server RPC port	bidirectional	All nodes	All nodes	Port used for search service internal com- munication
4243	RPC	Sage auto complete server metadata sub- scriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4244	RPC	Sage auto complete server metadata sub- scriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4245	RPC	Sage auto complete server metadata sub- scriber port	bidirectional	All nodes	All nodes	Port used for search internal communication
4249	TCP	Ports used by Enlite/ SearchIQ	bidirectional	All nodes	All nodes	Port used for SpotIQ internal communication
4251	RPC	Sage master RPC port	bidirectional	All nodes	All nodes	Port used for search service internal com- munication
4405	RPC	Diamond (graphite) port	bidirectional	All nodes	All nodes	Port used for communication with monitoring service

Port	Protocol	Service Name	Direction	Source	Dest.	Description
4406	RPC	Diamond (graphite) port	bidirectional	All nodes	All nodes	Port used for communication with monitoring service
4500	RPC	Trace vault service RPC port	bidirectional	All nodes	All nodes	Trace collection for ThoughtSpot services
4501	HTTP	Trace vault service HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug trace collection
4851	RPC	Graphite manager RPC port	bidirectional	All nodes	All nodes	Communication with graphite manager
4852	HTTP	Graphite manager HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug graphite manager
4853	RPC	Elastic search stack (ELK) manager RPC port	bidirectional	All nodes	All nodes	Communication with log search service
4853	HTTP	Elastic search stack (ELK) manager HTTP port	bidirectional	Admin IP addresses and all nodes	All nodes	Debug log search service
9200	RPC	Elastic search (ELK)	bidirectional	All nodes	All nodes	Communication with log search service
5021	RPC	Callosum services like meta-data services, metadata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts metadata service for metadata
5432	Postgres	Postgres database server port	bidirectional	All nodes	All nodes	Communication with Postgres database

Port	Protocol	Service Name	Direction	Source	Dest.	Description
6021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata
7021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata
8020	RPC	HDFS namenode server RPC port	bidirectional	All nodes	All nodes	Distributed file system (DFS) communication with clients
8021	RPC	Callosum services like meta-data services, medata-dependency service, scheduling service, session-less service, spotiq service	bidirectional	All nodes	All nodes	Port where search service contacts meta-data service for metadata
8080	HTTP	Tomcat	bidirectional	All nodes	All nodes	BI engine communication with clients
8081	HTTP	Callosum/Tomcat status	bidirectional	All nodes	All nodes	BI engine communication with clients
8787	HTTP	Periscope (UI) service HTTP port	bidirectional	All nodes	All nodes	Administration UI back end
8888	HTTP	HTTP proxy server (tinyproxy)	bidirectional	All nodes	All nodes	Reverse SSH tunnel
11211	Memcached	Memcached server port	bidirectional	All nodes	All nodes	BI engine cache
12345	ODBC	Simba server port	bidirectional	All nodes	All nodes	Port used for ETL (extract, transform, load)

Port	Protocol	Service Name	Direction	Source	Dest.	Description
8480	HTTP	HDFS journalnode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
8485	HTTP	HDFS journalnode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50070	HTTP	HDFS namenode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50090	HTTP	HDFS secondary namenode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS metadata
50075	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data
50010	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data
50020	HTTP	HDFS datanode server HTTP port	bidirectional	All nodes	All nodes	Debug DFS data
7000	TCP	Cassandra KV store database	bidirectional	All nodes	All nodes	Debug DFS data
7001	TCP	Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
9042	HTTP	Munshi server impression service, Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
9160	TCP	Cassandra	bidirectional	All nodes	All nodes	Debug DFS data
4010	HTTP	Falcon moderator	bidirectional	All nodes	All nodes	Debug DFS data
4011	HTTP	Falcon moderator	bidirectional	All nodes	All nodes	Debug DFS data

Port	Protocol	Service Name	Direction	Source	Dest.	Description
20123 - 32768	TCP (dynamic)	Dynamic port in this range used for various services and ancillary services like atlas, caffeine, call-home, callosum, falcon, monitoring, munshi server, nlp, object_search, postgres, sage UBR, spo-tiq snapshot, timely	All nodes	Services		
5270	TCP	Cluster monitoring service (ELK)	bidirectional	All nodes	All nodes	Services
5271	TCP	Cluster monitoring service (ELK)	bidirectional	All nodes	All nodes	Services
5601	TCP	Kibana UI (ELK)	bidirectional	All nodes	All nodes	Services
6311	TCP	R service	bidirectional	All nodes	All nodes	Services
8008	TCP	Video recorder	bidirectional	All nodes	All nodes	Services
9090	TCP	Timely	bidirectional	All nodes	All nodes	Services
ICMPv4		Used for health check of cluster nodes	bidirectional	All nodes	All nodes	Services

Ports for inbound and outbound cluster access

ThoughtSpot uses static ports for inbound and outbound access to the cluster.

Port	Protocol	Service Name	Direction	Source	Destination	Description
22	SCP	SSH	bidirectional	ThoughtSpot Support	All nodes	Secure shell access.

Port	Protocol	Service Name	Direction	Source	Destination	Description
25 or 587	SMTP	SMTP or Secure SMTP	outbound	All nodes and SMTP relay (provided by customer)	All nodes	Allow outbound access for the IP address of whichever email relay server is in use. This is for sending alerts to ThoughtSpot Support. In ThoughtSpot release 6.1.1 or later, and in release 6.0.5, you can specify a custom port to connect to the relay host, instead of port 25 or 587. Refer to Set the relay host for SMTP [See page 89] .
80	HTTP	HTTP	bidirectional	ThoughtSpot Support	All nodes	Hypertext Transfer Protocol for website traffic.
123	UDP	NTP service	bidirectional	ThoughtSpot Support	All nodes	Port used by NTP service.
389 or 636	TCP	LDAP or LDAPS	outbound	All nodes and LDAP server, provided by customer	All nodes	Allow outbound access for the IP address of the LDAP server in use.
443	HTTPS	HTTPS	bidirectional	ThoughtSpot Support	All nodes	Secure HTTP.
443	TCP	HTTPS	outbound	All nodes	208.83.110.20	For transferring files to thoughtspot.egnyte.com.
443	TCP	HTTPS	outbound	All nodes	For transferring product usage data to mixpanel cloud.	outbound
443	TCP	HTTPS	outbound	All nodes	je8b47jfif.execute-api.us-east-2.amazonaws.com s3.us-west-1.amazonaws.com s3-us-west-1.amazonaws.com s3.dualstack.us-west-1.amazonaws.com	For transferring monitoring data to InfluxCloud. (Given address will resolve to point to AWS instances).

Port	Protocol	Service Name	Direction	Source	Destination	Description
587 or 25	SMTP	SMTP or Secure SMTP	outbound	All nodes and SMTP relay (provided by customer)	All nodes	Allow outbound access for the IP address of whichever email relay server is in use. This is for sending alerts to ThoughtSpot Support. In ThoughtSpot release 6.1.1 or later, and in release 6.0.5, you can specify a custom port to connect to the relay host, instead of port 25 or 587. Refer to Set the relay host for SMTP [See page 89] .
636 or 389	TCP	LDAP or LDAPS	outbound	All nodes and LDAP server, provided by customer	All nodes	Allow outbound access for the IP address of the LDAP server in use.
2049	TCP	NFS: In case one needs to mount NFS share on TS node.	bidirectional	ThoughtSpot Support	All nodes	Port used by NFS.
9000	HTTP	DataFlow	bidirectional	All users' IP address	Node that runs DataFlow	Port used for accessing DataFlow
12345	TCP	Simba	bidirectional	ThoughtSpot Support	All nodes	Port used by ODBC and JDBC drivers when connecting to ThoughtSpot.

Ports for Intelligent Platform Management Interface (IPMI)

ThoughtSpot uses static ports for out-of-band IPMI communications between the cluster and ThoughtSpot support.

Network ports

Port	Protocol	Service Name	Direction	Source	Dest.	Description
80	HTTP	HTTP	bidirectional	ThoughtSpot Support	All nodes	Hypertext Transfer Protocol for website traffic.
443	TCP	S-HTTP	bidirectional	ThoughtSpot Support	All nodes	IPMI GUI and for HTML5-based IPMI console access.
623	UDP	Serial-over-LAN	bidirectional	ThoughtSpot Support	All nodes	IPMI GUI and for HTML5-based IPMI console access.

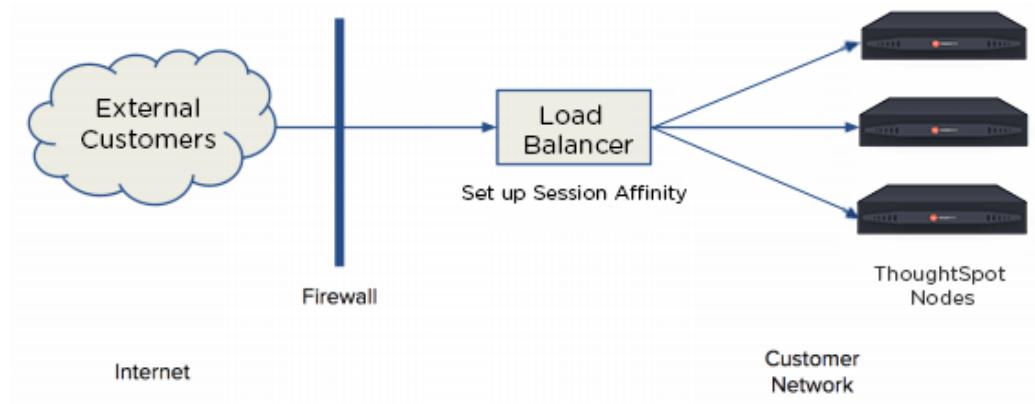
Configure load balancing and proxies

Summary: A load balancer is needed in front of a server group in order to direct traffic to individual servers in a way that maximizes efficiency.

Here are some of the best practices and guidelines for a typical implementation with ThoughtSpot. Your experience may differ depending on your environment and preference.

Load balance across ThoughtSpot nodes

The following shows a network architectural diagram which includes a load balancer for ThoughtSpot nodes.



The load balancer is an appliance in your infrastructure that routes traffic automatically to nodes to provide failover. You can also place a load balancer or proxy in front of the ThoughtSpot appliance if you'd like external network users to access the system.

The best way to load balance across all ThoughtSpot nodes in a cluster is to map one domain name (FQDN) to all the IPs in the cluster in a round robin fashion.

For example, if you want to use a DNS server based load balancing, then you can define multiple “A” resource records (RR) for the same name.

Below is an example of how you could set that up

```
thoughtspot.customer.com IN A 69.9.64.11  
thoughtspot.customer.com IN A 69.9.64.12  
thoughtspot.customer.com IN A 69.9.64.13  
thoughtspot.customer.com IN A 69.9.64.14
```

The example indicates that IP addresses for the domain thoughtspot.customer.com are 69.9.64.11, 69.9.64.12, 69.9.64.13, and 69.9.64.14.

Session Affinity

Session Affinity refers to directing requests to the same application server for the time it takes to complete a task.

In order for session affinity to work on ThoughtSpot, HTTPS (an SSL certificate) has to be installed on the load balancer level. If it is installed outside of the load balancer, session affinity may not occur and the ThoughtSpot system will fail.

Web proxies

You can access ThoughtSpot through any standard web proxy server. Web proxies are fairly universal regardless of the application they are proxying. However, ThoughtSpot doesn't use any new protocols, like SPDY or HTTP/2, which may have a dependency on the proxy. Instead, ThoughtSpot is commonly placed behind a web HTTP/HTTPS proxy.

Additionally, the proxy can round robin across multiple nodes in the ThoughtSpot backend. You can essentially use the web proxy as a load balancer. Therefore, your session will carry over if the proxy round robins between the ThoughtSpot backends as long as the URL doesn't change.

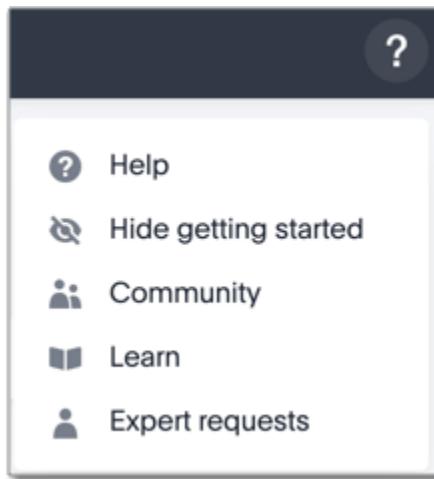
Customize ThoughtSpot Help

Summary: You can customize ThoughtSpot Help to be specific to your data, examples, and documentation.

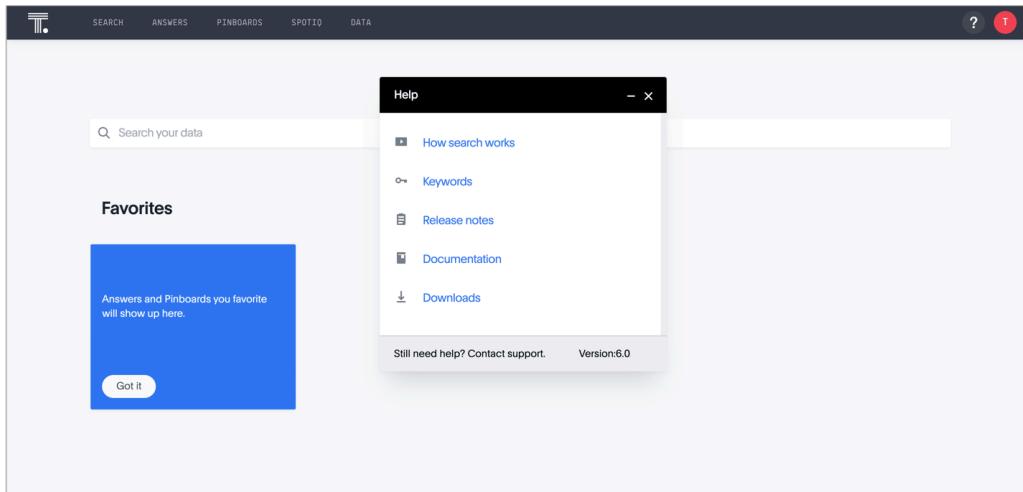
You can customize the Help menu for your ThoughtSpot application to tailor it to your organization.

When you configure these Help settings, you set system-wide defaults for all your users.

When your ThoughtSpot users click the Help icon, they see a list of links.

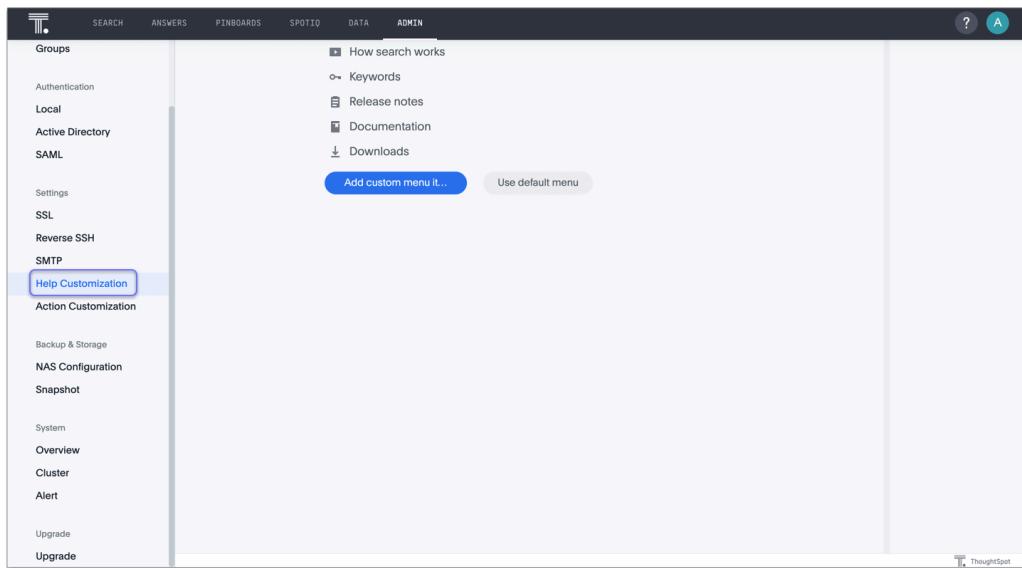


If a user clicks on the **help** link, the customizable help menu opens:

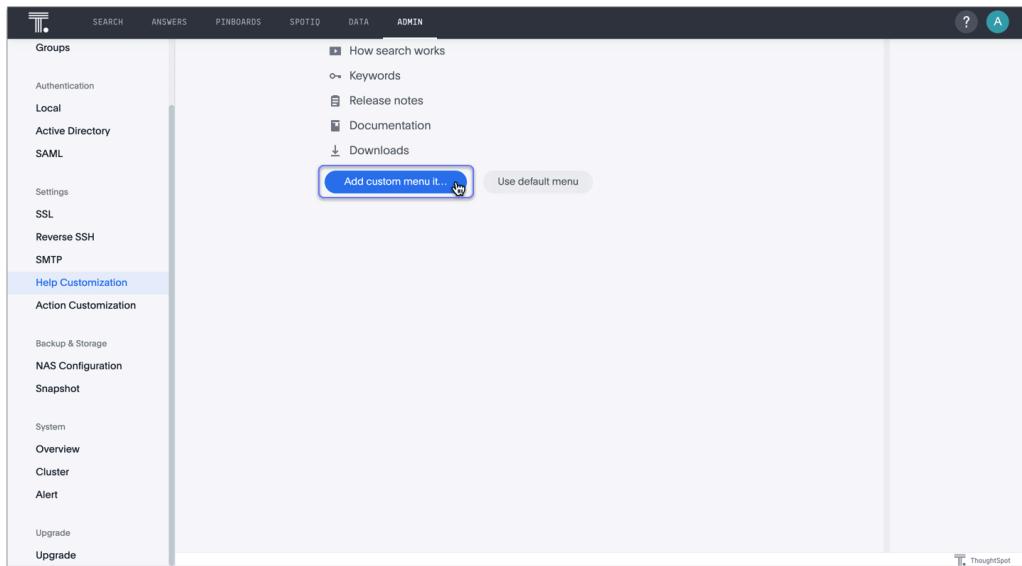


As an administrator, you can add your own links to this default list. This allows you to include documentation specific to your company, such as information about the data available in ThoughtSpot, where to get support internally, or company-specific training.

To customize the Help menu, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Help Customization** from the side navigation bar that appears.



To add a new link, select **Add custom menu item** below the list of existing links.



Specify the name of the link, the URL, and optionally add a custom icon. Then click **Confirm**.

Help menu customization

Item Label

URL

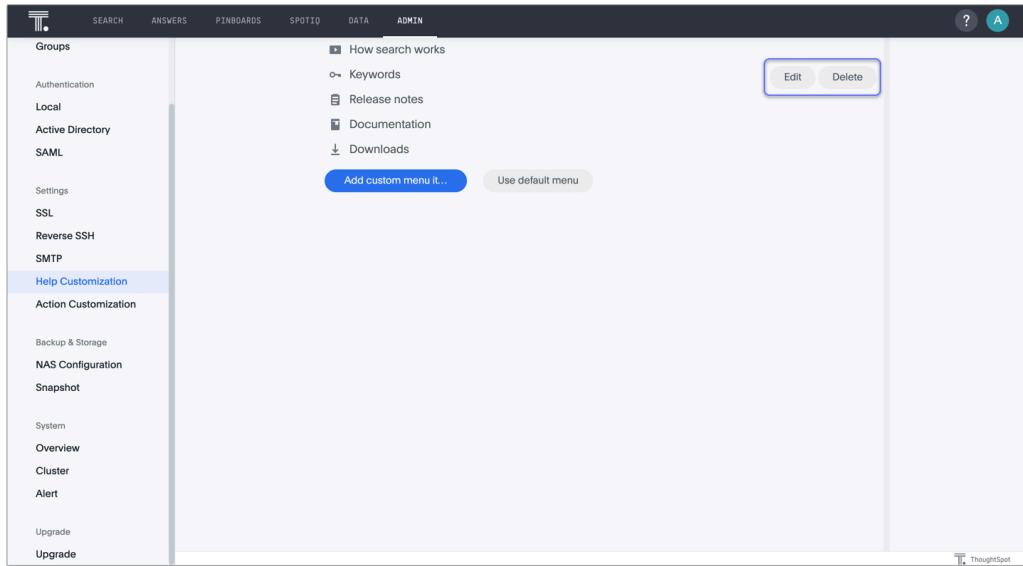
Custom Icon 

Item Enabled

Cancel **Confirm**

You can also edit existing links, change icons, or remove items from the help menu altogether.

To edit or delete a link, hover over the menu item and click the **Edit** or **Delete** button.



The screenshot shows the ThoughtSpot Admin interface with the 'Help Customization' section selected in the sidebar. The main content area lists several help menu items with their corresponding icons and descriptions. The 'Edit' and 'Delete' buttons are highlighted with a blue border, indicating they are interactive elements.

Specify the new information, and click **Confirm**.

Help menu customization

Item Label

URL

Custom Icon 

Item Enabled

Cancel Confirm

Customize look and feel

Summary: You can brand the ThoughtSpot application for your company.

You can customize the look and feel of the ThoughtSpot application for your company. Configuring these settings sets system-wide defaults for all your users.

Style customization is enabled by default. You can configure your cluster to disable this functionality. Contact [ThoughtSpot Support](https://community.thoughtspot.com/customers/s/contactsupport) (<https://community.thoughtspot.com/customers/s/contactsupport>) about disabling this feature.

Where to customize styles

A user must have administrative privileges.

To access style customization,

1. Click **Admin** on main navigation.
2. Under **Settings**, select **Style Customization**.

General guidelines for customization

Your changes take effect either immediately, or after you refresh the browser. You can revert your changes by using the **Reset** button that appears when you move your cursor moves to the right of the setting.



You can configure these style customizations:

Application Logo (Default) & Favicon

Sets a default application and favicon logo. This should be 140 pixels square.

Application Logo (Wide)

This logo should be 440 x 100 pixels.

Chart Visualization Fonts

Set a font for chart labels. You can specify any Web Open Font Format (WOFF) file.

Table Visualization Fonts

Set a font for table labels. You can specify any Web Open Font Format (WOFF) file.

Embedded Application Background

Set the background for an embedded ThoughtSpot instance. This is only used if you are embedding ThoughtSpot in another application.

Chart Color Palettes

Set the default palette for all charts. To set a value, however over a color value and enter a HEX value or select one from the chart. Individual users can still customize their own chart colors. They can use the **Reset colors** option on a chart to clear their changes.

Footer text

Define a footer to appear with the ThoughtSpot application.

Page title

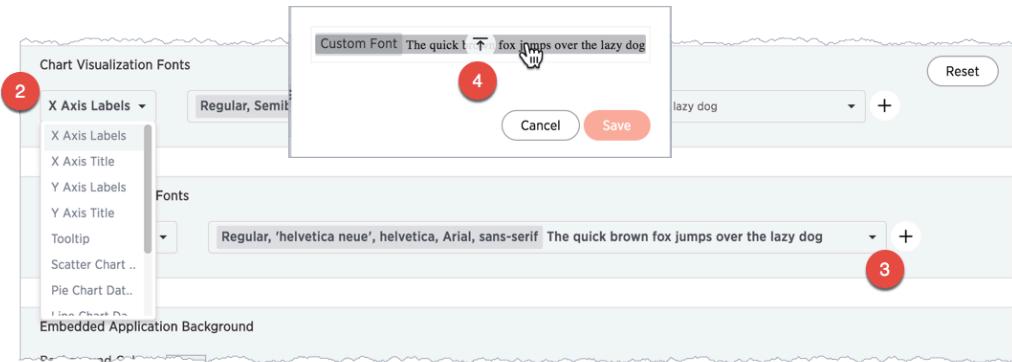
Sets the title for the browser tab.

How to specify fonts

You can set your systems default fonts by specifying either or both of the **Chart Visualization Fonts** and **Table Visualization Fonts** values. To set this values, your font must be defined in a Web Open Font Format (WOFF) file.

Changes to chart and table defaults apply only to charts and features created *after* you configure a value. If you change table fonts, older tables retain their previous fonts. Some settings can be overridden on a per-object level. Users cannot configure their own defaults.

To set a new font:



1. Make sure you have a WOFF file available for your font.
2. Select the label you want to change.
3. Click the + button.

The system displays the **Custom Font** dialog.

4. Click the **Custom Font** field.

The system displays the file finder.

5. Add the `WOFF` file you want.

The file appears in the font dialog

6. Click **Save** to change the font.

How to specify the behavior or clickable links in data

There is a system-wide setting which determines what happens when a user clicks on a link within the data. When your data includes URLs, they display as clickable links in ThoughtSpot tables. By default, clicking on a link opens the URL in a separate tab. But there is a system-wide setting that can be changed to open the links within the context of where they appear.

Consider how the links open, by type:

Link in search result table in ThoughtSpot

Same browser tab as ThoughtSpot application

Link in table embedded in an <iframe>

Same <iframe> that contains the table

Link in full ThoughtSpot application embedded in an <iframe>

Same <iframe> that contains the application

Load and manage data

Summary: Learn about loading and managing data.

The fastest and easiest way to load a new table is by importing it using the Web browser. This is best for one time data loads of small tables which do not have complex relationships to other tables. This method is limited to tables that are under 50 MB (megabytes) in size.

Using ThoughtSpot Loader, you can script recurring loads and work with multi-table schemas.

If your data already exists in another database with the schema you want to use in ThoughtSpot, you can pull the schema and data in using the ODBC or JDBC driver.

These are the methods you can use to load data, along with the benefits of each method:

Method	Description	Benefits
Connect to external databases (See https://docs.thoughtspot.com/6.0/data-integrate/embrace/embrace-intro.html)	Use the ThoughtSpot Embrace to read directly from the external databases.	Easy way to set up and enable the connection between ThoughtSpot and external databases. Users can send live query to the external databases, without having to replicate data again in ThoughtSpot for analysis.
Load data from the ThoughtSpot UI [See page 173]	Use the ThoughtSpot Web interface to upload an Excel or CSV (comma separated values) file from your local machine.	Easy way to do a one-time data load of a small file (under 50MB). End users can upload their own data and explore it quickly.
Import with the ThoughtSpot Loader (tsload) [See page 238]	Use TSQL and tsload to load data directly into the back end database that ThoughtSpot uses.	Best way to load large amounts of data or a schema with multiple tables. Can be scripted and used for recurring data loads, such as monthly sales results or daily logs. Can be integrated with an ETL solution for automation.
Use the ODBC/JDBC driver to connect to ThoughtSpot	Use the ODBC or JDBC client with your ETL tool. For information, see the ThoughtSpot Data Integration Guide.	Make use of an established ETL process and tool(s). Connect to ThoughtSpot using third party tools like SSIS. You don't need to define a schema to accept the data load.

If you're uploading data through the Web interface, you can use a native Excel file. If you want to use a CSV (comma separated values) or delimited file, or you are loading using ThoughtSpot Loader, you must [create CSV files with the data to be loaded \[See page 174\]](#) first.

Tip: End users will almost always work with worksheets and data they upload.

Related Information

- [Load CSV files with the UI \[See page 173\]](#)
- [Append data through the UI \[See page 0\]](#)
- [Schema planning concepts \[See page 184\]](#)
- [Overview of schema building \[See page 211\]](#)
- [Import CSV files with tsload \[See page 238\]](#)
- [How to view a data schema \[See page 178\]](#)

Configure casing

Summary: You can set the type of case sensitivity you would like to see reflected in the ThoughtSpot display.

Before you load your data, you should consider the type of casing you would like your data to reflect.

The case sensitivity for source data strings is preserved in the display. So, the visual display of results is identical to the input case that is loaded.

Note: The casing will remain lowercase in other parts of the application, such as when you ask a question or filter.

It is important to note that string casings aren't applied globally, but by column. So datasets will have different string casings as long as they're in different columns. Tables that are already compacted will keep their lowercase format. In these cases, to get the specific string case that you want, you would have to truncate related tables and reload them.

To take advantage of case configuration, you need to have ThoughtSpot Support enable it on your cluster for you. In addition, title casing should be disabled for string casing to properly work.

Load CSV files with the UI

Summary: The simplest way to load data is to upload a CSV or Excel file from the ThoughtSpot Web interface.

Loading data through the Web browser is recommended for smaller tables (under 50MB) with simple relationships between them. This method is recommended for small, one time data loads. Using this method, the data schema is created for you automatically.

Any user who belongs to a group that has the privilege **Has administration privileges** or **Can upload user data** can upload their own data from the browser.

Your data should be in a CSV (comma separated values) before you load it. A CSV file is a text file made up of data fields separated by a delimiter and optionally enclosed with an enclosing character. If your data contains multiple tables, you can have a separate CSV for each table.

Formatting the CSV

Your ETL (extract, transform, load) process will typically generate CSV files. You can also create a CSV file from a Microsoft Excel spreadsheet by opening the spreadsheet in Excel, choosing **Save As** and selecting CSV.

A CSV file contains a delimiter that marks the separation between fields in the data. The delimiter is usually comma, but it can be any character. The file also contains fields optionally enclosed with double quotes. Use these guidelines when creating the CSV file:

- If the CSV contains column headers, they must match the column names in the database exactly.
- Often a `|` (pipe) or tab is used as the delimiter, because it may be less likely to occur within the data values.
- When a field contains a double quote, it must be escaped with the character specified in the escape character argument in `ts load`.
- When a field contains the delimiter, the field must be enclosed in double quotes.

ThoughtSpot supports a wide range of date and timestamp formats (See <https://docs.thoughtspot.com/6.0/reference/date-formats-for-loading.html#>) in the CSV file. Blank values in user uploaded CSV files are interpreted as NULL values. These include the values (case insensitive):

- `NULL`
- `\N`
- `NA`
- `N/A`
- `[space]`

If you are appending data to an existing schema or table, columns in the CSV file must be in the same order as defined in the target table.

If you are loading a fact table that joins to dimension tables, you must load the fact table first, and then the dimension tables. The joining key must be a single column of unique values in the dimension table. `NULL` values in the fact table cannot be joined.

Create a CSV file

The first step in loading data is to obtain or create one or more CSV files that contain the data to be

loaded into ThoughtSpot. CSV is a common format for transferring data between databases.

ThoughtSpot requires this format.

Most applications such as Microsoft Excel or Google Sheets can output CSV formatted files. If your source is an Excel spreadsheet or Google Sheet:

1. Save, export, or download the file in CSV format. The exact procedure you use will depend on the source application.
2. Review the file's format before uploading it to ThoughtSpot.

Your source data may be in another database. If this is the case, your company's ETL (extract, transform, load) process will typically generate CSV files. If your source is another database:

3. Connect to the source database.
4. Extract each table you wish to import into ThoughtSpot as a CSV file.

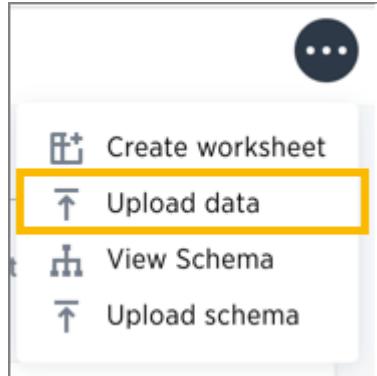
The column delimiter should be a **,** (comma), **|** (pipe), or tab.

For general information about CSV files and the rules for creating them, see the Comma-separated_values on Wikipedia (http://en.wikipedia.org/wiki/Comma-separated_values).

Load the CSV File

Any user who belongs to a group that has the privilege **Has administration privileges** or **Can upload user data** can upload their own data from the browser. To load the CSV or Excel file into ThoughtSpot:

1. Log in to ThoughtSpot from a browser.
2. Click **Data**, on the top navigation bar.
3. Click the ellipses icon  , in the upper right corner, and select **Upload Data**.



4. Upload the CSV or Excel file by doing one of these options:
 - a. Click **Browse your files** and select the file.
 - b. Drag and drop the file into the drop area.
5. Answer the question **Are the column names already defined in the file header?**
6. Answer the question **Are the fields separated by?** Click **Next**.

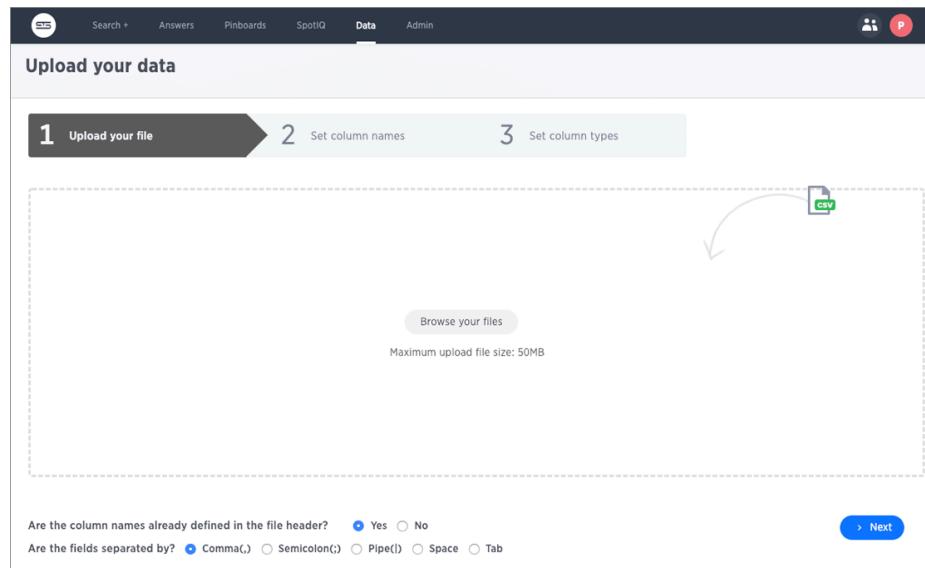
7. Click the column header names to change them to more useful names, if you'd like. Click **Next**.

8. Review the automatically generated data types for each column, and make any changes you want.

There are four data types: Text, Integer, Decimal, and Date.

9. Click **Import**.

When an upload is complete, the system reports the results and offers you some further actions.



- Click **Link to Existing Data** if you want to link the data you uploaded to the data in another table or worksheet.
- Click **Search** if you want to begin a new search.
- Click **Auto analyze** if you want to use the SpotIQ feature to find insights in your new data.

Append to an existing table

You can append data to your existing system tables through the ThoughtSpot application, even if the tables were initially loaded using `ts load`. The CSV file must have the same structure as the table it is being loaded into, including number and type of columns, in the same order as the target table.

To append data into ThoughtSpot:

1. Log in to ThoughtSpot from a browser.
2. Click **Data** on the top navigation bar.
3. Click the name of the table you would like to append data to.
4. Click the **Load Data** button.

The screenshot shows the ThoughtSpot Data interface. At the top, there's a navigation bar with links for Search +, Answers, Pinboards, SpotIQ, Data, Admin, and a user icon. Below the navigation bar, the title 'SYSTEM TABLE ThoughtSPORT_Product_Dimension' is displayed. Underneath the title, there's a horizontal menu with tabs: Columns (which is underlined), Schema, Data, Profile, Dependents, Row Security, and SpotIQ Insights. The main area shows a table with five rows of data. The columns are: COLUMN NAME, DESCRIPTION, DATA TYPE, COLUMN TYPE, ADDITIVE, AGGREGATION, HIDDEN, SYNONYMS, and INDEX TYPE. The data rows are: Product_Key, Click to edit, INT32, ATTRIBUTE, NO, NONE, NO, Click to edit, DONT_INDEX; Product_Name, Click to edit, VARCHAR, ATTRIBUTE, NO, NONE, NO, Click to edit, DEFAULT; SKU_Number, Click to edit, VARCHAR, ATTRIBUTE, NO, NONE, NO, Click to edit, DEFAULT; Department_Desc., Click to edit, VARCHAR, ATTRIBUTE, NO, NONE, NO, Click to edit, DEFAULT; Category, Click to edit, VARCHAR, ATTRIBUTE, NO, NONE, NO, Click to edit, DEFAULT. A note at the bottom left says '(showing rows 1-5 of 5)'. In the top right corner of the table area, there's a 'Load Data' button with three dots next to it, and the 'Load Data' button is highlighted with a red box.

5. Upload the CSV or Excel file by doing one of these options:
 - Click **Browse your files** and select the file.
 - Drag and drop the file into the drop area.
6. Answer the question **Are the column names already defined in the file header?**.
7. For the question **Do you want to append to the existing data or overwrite it?**, select **Append**.
8. Answer the question **Are the fields separated by?**, and click **Next**.
9. Click **Upload**.
10. Click **Link to existing data** if you want to link the data you uploaded to the data in another table or worksheet. Or click **Ask a question** if you want to begin a new search.

How to view a data schema

Summary: Use the schema viewer to see tables and worksheets and their relationships.

ThoughtSpot has a **Schema Viewer** through which you can examine the database schema. It is interactive and configurable, so you can see the level of detail that is relevant to your work.

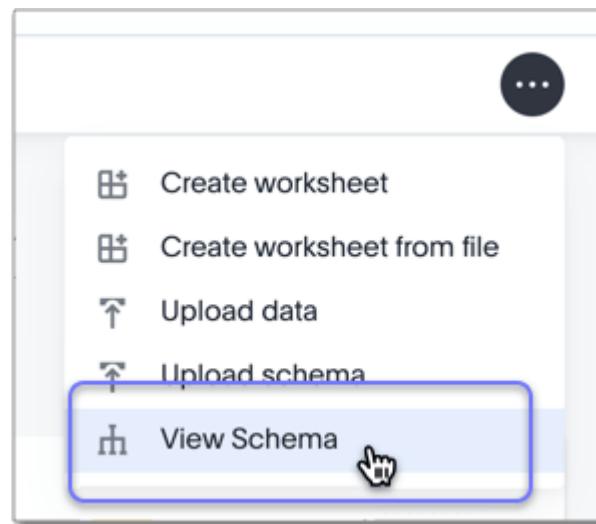
You must have **Admin** privileges to use the **Schema Viewer**.

Accessing the Schema Viewer for all objects

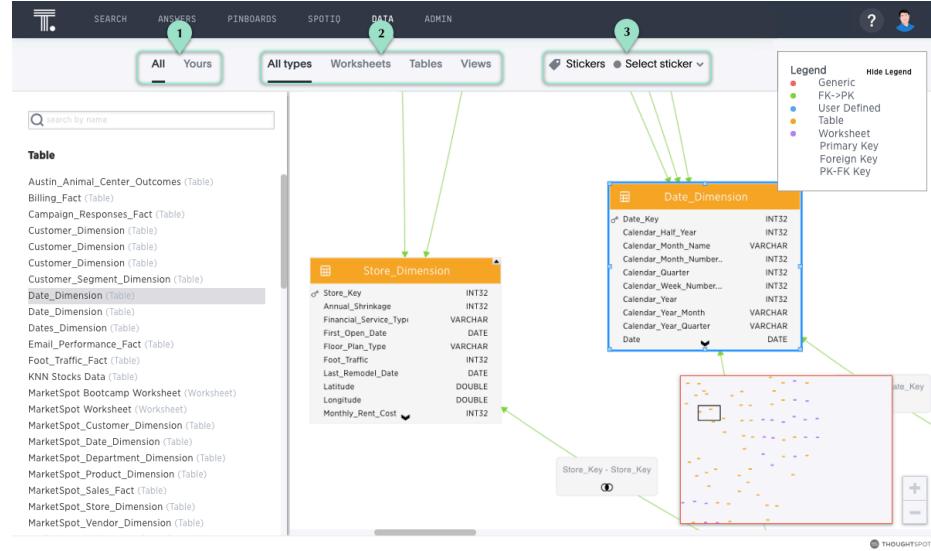
1. Click the **Data** tab in the top menu.



2. Click the **More** menu (ellipsis icon), and select **View Schema**.



3. When the schema appears, you can see that you can control the visible part of the schema.



Legend Action

1. See either All (Default), or Yours.
2. See either All types (Default), Worksheets, Tables, or Views.
3. Select artifacts tagged with stickers.
4. The list of tables, worksheets, and imported data on the left changes as you select the various filters. The schema view focus changes in tandem.
5. To center the view panel on a specific table, worksheet, or view, click that object.

You can also drag the objects around in the viewer to position them better.

Accessing Schema Viewer for a single worksheet, table, or view

You can now see the schema for each object for tables, worksheets, or views.

1. Click the **Data** tab in the top menu.



2. Select from the list of the possible objects:

Legend Action

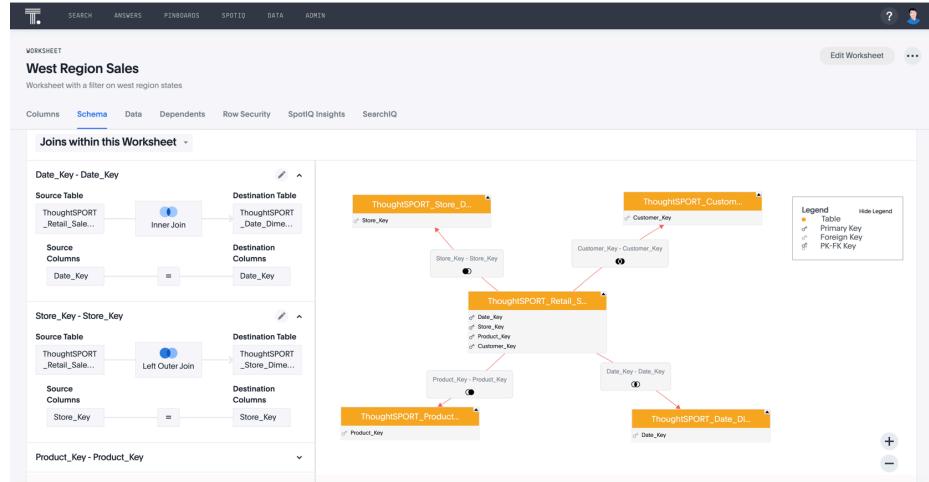
1. See either *All* (Default), or *Yours*.
2. See either *All types* (Default), *Worksheets*, *Tables*, or *Views*.
3. Select artifacts tagged with stickers.

3. From the list of objects, select one. Here, we selected the worksheet *West Regional Sales*.
4. At the top of the worksheet, click the **Joins** tab.

Legend Action

1. See either *All* (Default), or *Yours*.
2. See either *All types* (Default), *Worksheets*, *Tables*, or *Views*.
3. Select artifacts tagged with stickers.

5. The join information and the schema for the worksheet appear.



Why use the Schema Viewer

You can use the Schema Viewer to discover the following information:

- What is the relationship between two tables?
- What tables make up this worksheet, and how are they joined?

The schema viewer shows joins between tables, join directionality, and join type.

How the Schema Viewer shows joins

You can use the Schema Viewer to review your schema and ensure that it was modeled using best practices. For example, joins appear in different colors to distinguish their type:

- Generic relationships are in **red**
- Primary key/ foreign key joins are in **green**

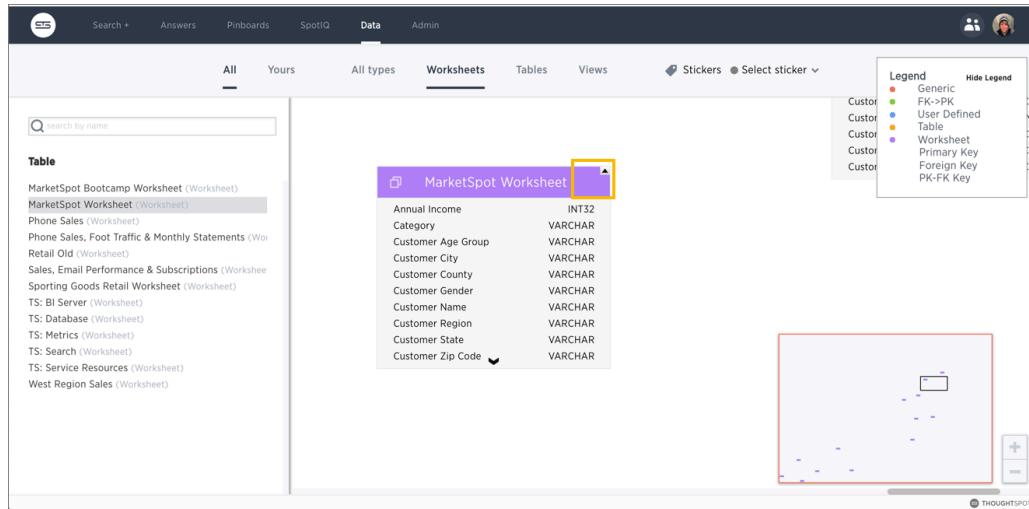
When viewing a worksheet, you can also see what joins connect the tables: the inner, left outer, right outer, or full outer joins.

Note: Defining a generic relationship in the UI rather than using a primary key/ foreign key join through TQL has no impact on performance. However, when creating relationships in the UI, you must ensure that you create it in the right direction: many to one. To create many-to-many joins, or to create joins using >, <, >=, or <=, use TQL.

Worksheet view

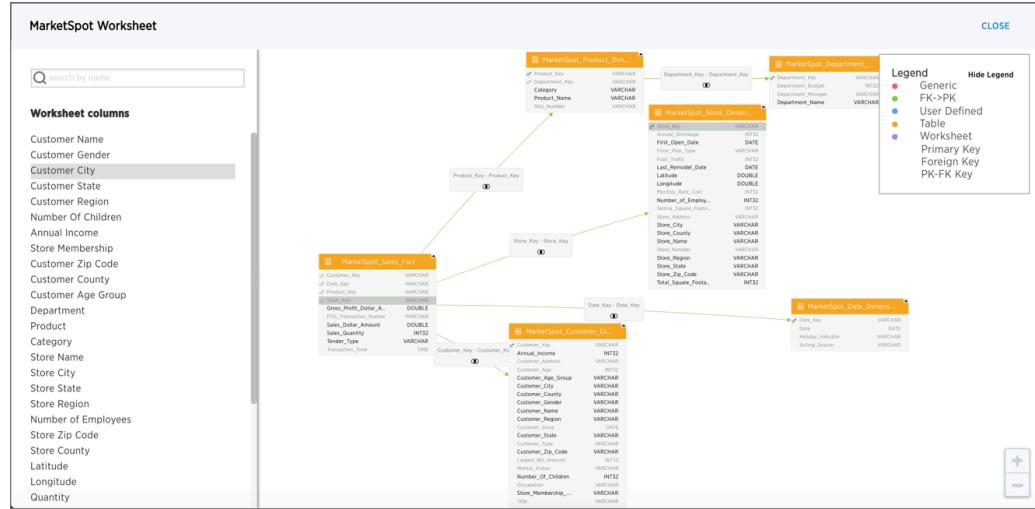
Worksheets are often based on more than one table. The worksheet schema will show schemas for the tables behind the worksheet, as well as the joins between tables *that were created as a part of the worksheet*.

Click a worksheet, to see it in the Schema Viewer. If the schema view is not showing the schema behind the worksheet, double click the tab on the top right of the worksheet object.



The worksheet view shows the following information:

- All tables in the worksheet, and the relationships between these tables.
- Source columns for all columns of a worksheet.
- Keys and definitions for each relationship, as well as join paths and types.
- Columns that are derived from formulas.
- Correct join paths for newly created chasm trap worksheets. Chasm trap worksheets created prior to ThoughtSpot version 4.4 do not show the correct join paths.



Related Information

- [Worksheet joins \[See page 483\]](#)
- [Modify joins within a worksheet \[See page 488\]](#)
- [Change the schema using TQL \[See page 228\]](#)
- [Constraints \[See page 191\]](#)

Schema planning concepts

Summary: Learn about creating a schema for the ThoughtSpot Loader.

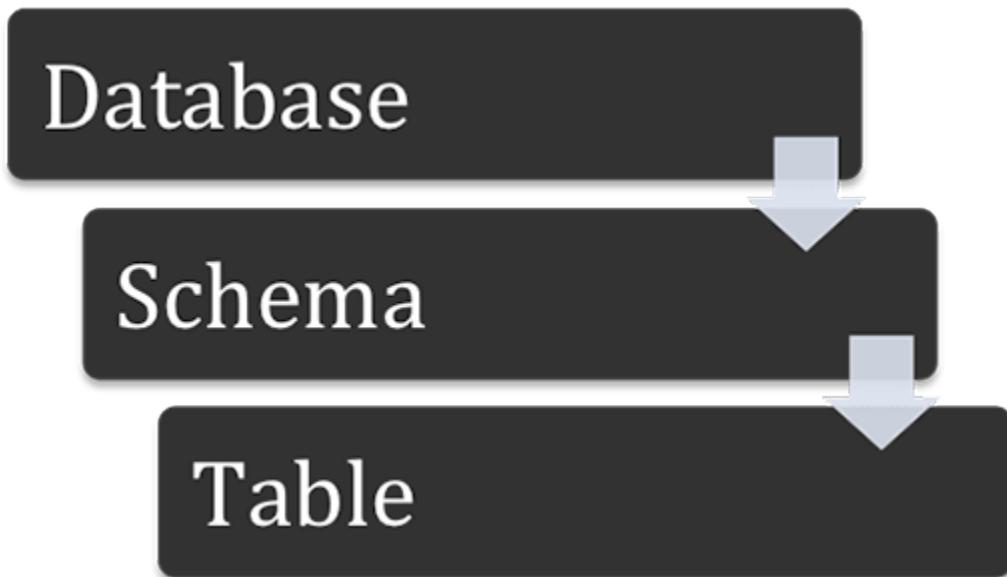
Before you can load data with ThoughtSpot Loader, you must create a schema to receive it, using the SQL command line interface (TQL).

The TQL syntax is similar to the SQL used in other relational databases, but with some important differences. Use DDL (data definition language) to create the schema into which you can load the data. We recommend combining all DDL statements into a single script for creating the schema.

Before writing your TQL script, you need to understand some basic ThoughtSpot concepts.

About databases and schemas

ThoughtSpot organizes objects in a hierarchical namespace. Databases contain schemas, which contain tables.



ThoughtSpot can contain one or more databases, and each database can have multiple schemas. If you do not specify a schema, the default schema (`falcon_default_schema`) is used automatically. This makes it easier to add tables to the database without the need to explicitly create a schema.

If you do create an additional schema, you must refer to its objects using the syntax

`< schemaname>.< objectname >` . If you do not qualify the schema when referencing its objects, the default schema (`falcon_default_schema`) will always be assumed.

By default, ThoughtSpot creates an internal database to host tables corresponding to data that is imported by users from a Web browser.

Review the structure of your data

The schema you create to hold the data needs to be a good fit for your data. First, familiarize yourself with the tables you want to load, and understand their structure. Make note of this information for each table:

- The column names and data types
- Type of table (fact or dimension)
- Primary key column(s)
- The size of the table on disk
- Any other tables it can be joined with (foreign keys)

Here's what you must take into account in your TSQL for creating each table, based on these properties:

Table type	Table size	To be joined with	Schema recommendations
Fact	Any	Small dimension table(s)	Sharded. Foreign key references the primary key in the dimension table.
Fact	Any	Large dimension table(s)	Sharded on the same distribution key as the dimension table it will be joined with. Foreign key references the primary key in the dimension table.
Fact	Any	Another fact table	Sharded on the same distribution key as the fact table it will join with. Many-to-many relationship defines how the tables will be joined.
Dimension	under 50MB	Fact table(s)	Replicated (not sharded). Has a primary key.

Table type	Table size	To be joined with	Schema recommendations
Dimension	over 50MB	Fact table(s)	Distributed dimension table, sharded on the same distribution key as the fact table it will be joined with. Primary key must be the same as the distribution key.

Where to go next

- [Data types \[See page 187\]](#)

ThoughtSpot supports the common data types. Compare these with the data types you want to load, and do any necessary conversion ahead of loading the data.

- [Constraints \[See page 191\]](#)

Constraints include primary keys, foreign keys, and relationships. Relationships allow you to create a generic relationship for use when you want to join tables that don't have a primary key/foreign key relationship.

- [Sharding \[See page 196\]](#)

For the best performance, you should split (or shard) very large tables across nodes. If you have a large dimension table, you might choose to co-shard it with the fact table it will be joined with.

- [Chasm traps \[See page 208\]](#)

In a complex schema, you may have a fact table with no relationship to another fact table, except that each contains a foreign key to a shared dimension table. This is known as a chasm trap, and ThoughtSpot can handle it!

Data types

Summary: ThoughtSpot supports the common data types.

Before you import data, compare the data types you want to load with these supported data types.

Then, convert your data before loading it. Typically, you would export the data, transform it to meet these type rules, and then load the data. This is known as an extract-transform-load process.

Supported data types

The tables you create to receive the data must have the same number of columns and data types as the data you will be loading. Choose a data type for each column from the list of supported data types:

Data	Supported data types	Details
Character	VARCHAR(*n*)	Specify the maximum number of characters, as in VARCHAR(255). The size limit is 64MB for VARCHAR values.
Floating point	DOUBLE or FLOAT	DOUBLE is recommended. DOUBLE has a range of 1.7E +/- 308 (15 digits).
Boolean	BOOL	Can be true or false.
Integer	INT or BIGINT	INT holds 32 bits. BIGINT holds 64 bits. INT has a range of -2,147,483,648 to 2,147,483,647. BIGINT has a range of -9,223,372,036,854,775,808 to 9,223,372,036,854,775,807.
Date or time	DATE , DATETIME , TIMESTAMP , TIME	DATETIME , TIMESTAMP , and TIME are stored at the granularity of seconds. TIMESTAMP is identical to DATETIME , but is included for syntax compatibility.

⚠ Warning: There is a 64MB limitation on the number of characters for VARCHAR. If you have any VARCHAR data that exceeds this limit, the entire load will fail.

Geographical data types

ThoughtSpot supports geographical data.

How to import geographical data

Import your geographical data as **text** values. This ensures that the data defaults to the correct configuration, where the data type is `VARCHAR`. You can use `DOUBLE` or `VARCHAR` for latitude and longitude data. See the following example:

COLUMN NAME	DATA TYPE	COLUMN TYPE	ADDITIONAL	AGGREGATION	HIDDEN	SYNONYMS	SPOTIQ PREFERENCE	INDEX TYPE	GEO CONFIG
Store Name	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to ...	DEFAULT	DEFAULT	None
Store State	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to ...	DEFAULT	DEFAULT	State
Store Region	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to ...	DEFAULT	DEFAULT	None
Store Zip Code	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to ...	DEFAULT	DEFAULT	Zip Code

In the above example, the zipcodes were imported as text values. By default, the `data type` is `VARCHAR`, the `column type` is `attribute`, and `additive` is `no`. You must specify the `geo config` for your data yourself. ThoughtSpot does not specify geo config automatically.

If you import zipcodes as numeric values, the `column type` defaults to `measure`. In ThoughtSpot, a `measure` is a numeric value that you can use in mathematical formulas. If you import your geographical data as numeric values, you **must** change the `column type` to `attribute` and specify `additive` as `no`.

Latitude and longitude

For latitude and longitude, you can use either `VARCHAR` or `DOUBLE`. Note that your latitude and longitude data must be in the form of positive and negative numbers, and not in the form of degrees. North of the Equator, latitude values are **positive**, and south of the Equator, latitude values are **negative**. East of the Prime Meridian, longitude values are **positive**, and West of the Prime Meridian, longitude values are **negative**.

Designate your geographical data in ThoughtSpot

After loading the data, designate it as a geographical data type when you [Edit the system-wide data model](#) [See page 408]. Wherever abbreviations or codes are used, they are the same as what the USPS (United States Postal Service) recognizes.

These data types can be designated as geographical data, which enables them to be visualized using the Geo chart types:

- Countries, for example:
 - United States
 - long_name : United States
 - name_sort : United States of America
 - abbreviation : U.S.A.
 - adm0_a3 : USA
 - adm0_a3_is : USA
 - adm0_a3_us : USA
 - admin : United States of America
 - brk_a3 : USA
 - brk_name : United States
 - formal_en : United States of America
 - iso_a2 : US
 - iso_a3 : USA
 - iso_n3 : 840
 - COUNTY for counties in the United States, for example:
 - santa clara county
 - pike county, ohio
 - pike county, OH
 - STATE_PROVINCE for states in the United States, for example:
 - name : California
 - US Postal Service abbreviation : CA
 - LATITUDE , which must be used with LONGITUDE , for example:

- 37.421023
- -1.282911
- `LONGITUDE`, which must be used with `LATITUDE`, for example:
 - 122.142103
 - -103.848865
- `ZIP_CODE` for zip codes in the United States, for example:
 - `po_name` : MT MEADOWS AREA
 - `ZIP` : "00012"
 - `zip2` : 12
- Other Sub-nation Regions, which are administrative regions found in countries other than the United States, for example:
 - bremen
 - normandy
 - west midlands

⚠ Important: You cannot upload your own custom boundaries.

Constraints

Summary: Constraints allow you to build relationships and join tables.

Constraints include primary keys, foreign keys, and relationships. Relationships allow you to create a generic relationship for use when you want to join tables that don't have a primary key/foreign key relationship.

Note: Defining a generic relationship in the UI rather than using a primary key/ foreign key join through TQL has no impact on performance. However, when creating relationships in the UI, you must ensure that you create it in the right direction: many to one. To create many-to-many joins, or to create joins using >, <, >=, or <=, use TQL.

Primary keys

When a primary key is selected for a table, it impacts data loading behavior. When a new row is added:

- If another row already exists with the same primary key, it is updated with the values in the new row.
- If a row with the same primary key does not exist already, the new row is inserted into the table.

This behavior is referred to as “upsert” because it does an `INSERT` or an `UPDATE`, depending on whether a row with the same primary key already exists.

Note that ThoughtSpot does not check for primary key violations across different shards of the table. Therefore, you need to shard the table on the primary key columns if you require this “upsert” behavior.

Permitted joins and necessary permissions

See this matrix for information about which joins you can create, and what permissions these joins require.

		Worksheets	View	Materialized View	Imported table (UI)	Table uploaded from backend (tsload) or through DataFlow	Table uploaded through Embrace	View on top of table uploaded through Embrace
	Necessary permissions:	None	None	None	Can manage data permission to load the table	Admin privileges to access tsload	None	None
Worksheets	Can edit permission on the source Worksheet	x	x	x	✓	x	x	x
View	Can edit permission on the source View	x	✓	✓	✓	✓	x	x
Materialized View	Can edit permission on the source Materialized View	x	✓	✓	✓	✓	x	x
						Note: It is a best practice to create this join through the UI, rather than using TQL .		
Imported table (UI)	Can edit permission on the source table	✓	✓	✓	✓	✓	x	x
Table uploaded from backend (tsload) or through DataFlow	Can edit permission on the source table	x	✓	✓	✓	✓	x	x
					Note: It is a best practice to create this join through the UI, rather than using TQL .	Note: This join is often used to create Worksheets.		

Table up-loaded through Embrace	Can edit permission on the source table, and can manage data permission	X	X	X	X	X	✓	Note: The two tables must be from the same connection.	✓	Note: The View and the table must be from the same connection.
View on top of table up-loaded through Embrace	Can edit permission on the source View	X	X	X	X	X	✓	Note: The View and the table must be from the same connection.	✓	Note: The two Views must be from the same connection.

Foreign key relationships

Foreign key relationships tell ThoughtSpot how two tables can be joined. These relationships are only used for joining the tables, and not for referential integrity constraint checking.

The directionality of primary key - foreign key relationships is important. The foreign key relationship is defined on the fact table and references the primary key(s) in the dimension table. So you can think of the fact table as the source and the dimension table as the target. In the schema viewer, notice that the arrow that represents a PK/FK join points to the dimension table.

If you use primary and foreign keys, when users search the data from the search bar, tables are automatically joined. For example, assume there are two tables:

- revenue, which is a fact table
- region, which is a dimension table

There is a foreign key on the fact table on `regionid` which points to the id in the region dimension table. When a user types in “revenue by region”, the two tables will be joined automatically.

Foreign keys have to match the primary key of the target table they refer to. So if there are multiple columns that make up the primary key in the target table, the foreign key must include all of them, and in the same order.

Generic relationships (many-to-many)

You may have a schema where there is a fact table that you want to join with another fact table. If there isn't a primary key/foreign key relationship between the tables, you can use many-to-many to enable this. You can do this by using the RELATIONSHIP syntax to add a link between them, that works similarly to the WHERE clause in a SQL join clause.

Note: Using many-to-many joins is not a best practice. In cases where you have two fact tables you want to join, it is better to find a way to create a bridge table between them, so you have a chasm trap. Look at your two fact tables to see if they share some common data that you could use to create a dimension table between them. For example, a date or product dimension could join an inventory fact table to a sales fact table.

Note: A many-to-many implementation does not protect from over counting in some searches. If you plan to use it, make sure your searches don't include aggregation or count searches that will count one value multiple times, because it satisfies the join condition for multiple rows.

This is a special kind of relationship, that applies to specific data models and use cases. For example, suppose you have a table that shows wholesale purchases of fruits, and another table that shows retail fruit sales made, but no inventory information. In this case, it would be of some use to see the wholesale purchases that led to sales, but you don't have the data to track a single apple from wholesale purchase through to sale to a customer.

In a many-to-many relationship, the value(s) in a table can be used to join to a second table, using an equality condition (required) and one or more range conditions (optional). These conditions act like the WHERE clause in a SQL JOIN clause. They are applied using AND logic, such that all conditions must be met for a row to be included.

To use a many-to-many relationship, you need to follow a few rules:

- There must be one equality condition defined between the two tables.
- Each table must be sharded on the same key that will be used for the equality condition.
- There can optionally be one or more range conditions defined.

This example shows the TQL statements that create the two fact tables and the relationship between them.

```
TQL> CREATE TABLE "wholesale_buys" (
    "order_number" VARCHAR(255),
    "date_ordered" DATE,
    "expiration_date" DATE,
    "supplier" VARCHAR(255),
    "fruit" VARCHAR(255),
    "quantity" VARCHAR(255),
    "unit_price" DOUBLE
) PARTITION BY HASH (96) KEY ("fruit");

TQL> CREATE TABLE "retail_sales" (
    "date_sold" DATE,
    "location" VARCHAR(255),
    "vendor" VARCHAR(255),
    "fruit" VARCHAR(255),
    "quantity" VARCHAR(255),
    "sell_price" DOUBLE
) PARTITION BY HASH (96) KEY ("fruit");

TQL> ALTER TABLE "wholesale_buys" ADD RELATIONSHIP WITH "retail_sales" AS "wholesale_buys"."fruit" = "retail_sales"."fruit" and ("wholesale_buys"."date_ordered" < "retail_sales"."date_sold" and "retail_sales"."date_sold" < "wholesale_buys"."expiration_date");
```

Sharding

Summary: Sharding partitions very large tables into smaller, faster, more easily managed parts called data shards.

ThoughtSpot tables can be replicated or sharded. Replicated tables exist in their entirety, as the complete data set, on each node. Sharded tables consist of a single data set, divided into multiple tables, or *shards*. The shards have identical schemas, but different sets of data.

Note that sharding and loading tables into ThoughtSpot only apply to ThoughtSpot's in-memory database. If you use [Embrace \[See page 0\]](#), and store your data in another data warehouse such as Snowflake or Amazon Redshift, your data modeling is done in that data warehouse, and not in ThoughtSpot.

When to shard your data

By default, ThoughtSpot replicates tables. To shard tables, you must add the `PARTITION BY HASH ()` clause to your `CREATE TABLE` statement.

Sharding your tables impacts the total amount of memory used by the table, its performance, and table loading times.

For example, you might shard a large table of sales data. You could divide a single sales table into shards that each contain only the data for a single year. The system then distributes these shards across several nodes. Requests for sales data are dispersed both by the year and the location of the shard in the node cluster. No single table or node is overloaded, improving both query performance and system load.

To optimize ThoughtSpot performance and memory usage, you should shard very large fact tables whenever possible. If you have a large dimension table, more than 40 million rows, you might choose to shard it along with the fact table it is joined with. Sharding both the fact and dimension table(s) is known as *co-sharding*. Refer to [Sharded dimension tables \[See page 204\]](#). Before co-sharding the fact and dimension tables, consult with your ThoughtSpot contact.

Table sizes and sharding recommendations

When you are considering which tables you need to shard, and how many shards to use, there are several key sizing guidelines to keep in mind. You may not be able to fulfill each of these requirements for every table you shard. The most important requirement is that the number of shards for a given table take up less than or equal to 60% of your total available CPU.

1. **60% of CPU:** The number of shards for a given table should not take up more than 60% of your total available CPU. When possible, the number of shards should be well under 60%, for queries that involve multiple sharded fact tables.

Note: 60% of CPU is a best practice. If necessary, you can use up to 80% of your available CPU. Consult with your ThoughtSpot contact before choosing a sharding strategy that requires 80% of your available CPU for a given table.

2. **Number of rows per shard:** Ideally, each shard should host about 15-20 million rows of data. So, if your table is under 20 million rows, you do not need to shard it. You should not have more than about 20 million rows of data on each shard.
3. **Ideal number of shards:** The ideal number of shards is the number of rows in a table divided by 20 million. So, if your table has fewer than 20 million rows, you do not need to shard it. However, if the ideal number of shards is more than 60% of your available CPU, you should ensure that you use less than 60% of your available CPU, rather than having the ideal number of shards.
4. **Maximum number of shards:** As a best practice, a given table should not have more than 1000 shards. This applies even if 60% of your CPU is above 1000. If you feel that you need more than 1000 shards for a given table, consult with your ThoughtSpot contact.
5. **The number of shards should be a multiple of the number of nodes:** To ensure equal distribution of data across all nodes, so that none of your nodes sits idle, the number of shards should be a multiple of the number of nodes. So, for a 12-node cluster, for example, a table could have 12, 24, 36, or 48 shards, and so on.

6. **Minimum number of shards:** Because the number of shards should be a multiple of the number of nodes, the minimum number of shards is the number of nodes. For a 12-node cluster, you should not have fewer than 12 shards.

This requirement may be difficult to achieve on large clusters with a high number of nodes. For example, you may have a table with 200 million rows on a 24 node cluster. Based on the guideline of 20 million rows per shard, this table should have 10 shards. 10 is not a multiple of 24. However, you may also have several very large tables on this cluster, with more than 1 billion rows. These 1 billion row tables can have at least 24 shards while fulfilling the 20 million rows per shard requirement, but the 200 million row table cannot. If you do not have these very large tables in your cluster, but you do have a high number of nodes, you might choose to have fewer than 20 million rows per shard, to ensure equal distribution of data. Consult with your ThoughtSpot contact if you are unsure how to handle sharding on your large cluster.

Sharding recommendations example

Let's use an example to see how the 6 guidelines listed above in [Table sizes and sharding recommendations \[See page 197\]](#) work.

You have a fact table with 2.4 billion rows. Your cluster has 24 nodes, and 56 CPU cores per node.

1. **Determine the ideal number of shards:**

Number of rows/20 million

$$2.4 \text{ billion}/20 \text{ million} = 120$$

2. **Determine 60% of the number of available CPU cores:**

*Number of nodes*Number of CPU cores per node*.6*

$$24*56*.6 = 806.4$$

3. **Compare the ideal number of shards with 60% of the number of available CPU cores:**

120 is less than 806.4, so you have enough CPU cores to shard this table with the ideal number of shards.

4. **Ensure that the number of shards is a multiple of the number of nodes, and therefore higher than the minimum number of shards for a given table:**

Number of rows/ Number of shards must equal a whole number to allow you to distribute shards equally among the nodes.

$120/24 = 5$. 5 is a whole number.

5. **Ensure that the number of shards is under 1000:**

120 is under 1000.

In summary: You can shard this table by the ideal number of shards; 120, in this case, because 120 is less than 60% of your total available CPU, it is a multiple of the number of rows, and it is under 1000.

If the table had 2.6 billion rows, the ideal number of shards would be 130. However, 130 is not a multiple of 24, so 120 would still be the best number of shards for this table.

How to shard

Sharding is a type of partitioning. It is sometimes called *Horizontal partitioning*. The term *sharding* is particular to situations where data is distributed not only among tables, but also across nodes in the system. To create a sharded table, add the `PARTITION BY HASH ()` clause to your `CREATE TABLE` statement.

```
TQL> CREATE TABLE
...
PARTITION BY HASH (96) KEY ("customer_id");
```

Note the following parameters, specified above as `96` and `"customer_id"`:

HASH

Determines the number of shards. In this case, `96`.

KEY

Specifies how to assign data into the shards (shard key). In this case, `customer_id`.

The recommended number of shards depends on various factors. See [Table sizes and sharding recommendations \[See page 197\]](#).

If you omit the `PARTITION BY HASH` statement or if the `HASH` parameter is 1 (one), the table is unsharded. The table is replicated instead, and physically exists in its entirety on each node. This increases memory usage, since you are storing multiple copies of the same table.

If you want to use a table's primary key as the shard key, specify that the table is to be partitioned by `HASH` on the primary key, as in this example:

```
TQL> CREATE TABLE "supplier" (
    "s_suppkey" BIGINT,
    "s_name" VARCHAR(255),
    "s_address" VARCHAR(255),
    "s_city" VARCHAR(255),
    "s_phone" VARCHAR(255),
    CONSTRAINT PRIMARY KEY ("s_suppkey")
) PARTITION BY HASH (96) KEY ("s_suppkey");
```

How to choose a shard key

Tip: We recommend that you always specify the `KEY` parameter when `HASH` is greater than 1.

If you omit the `KEY` parameter in your `CREATE TABLE` statement, ThoughtSpot shards the table randomly.

ThoughtSpot does not have a default shard key.

- If the table has no primary key, the sharding is unconstrained. You can choose **any** subset of columns that is valid for use as the primary key as the shard key. If you do not specify the shard key, ThoughtSpot implements random sharding.
- If the table has a primary key, you **must** specify the `KEY` parameter of the `PARTITION BY HASH` statement. This shard key **must** be a subset of the primary key.

DO

```
...
CONSTRAINT PRIMARY KEY("saleid, vendorid")
PARTITION BY HASH(n) KEY ("saleid");
```

```
...
CONSTRAINT PRIMARY KEY("saleid, vendorid")
PARTITION BY HASH(n) KEY ("vendorid");
```

In the above examples, the table has a primary key. The `KEY` parameters specified, `saleid` and `vendorid`, are subsets of the primary key.

In the below example, the table has a primary key. The `KEY` parameter specified, `locationid`, is *not* a subset of the primary key, and therefore cannot be used as the shard key.

AVOID

```
...
CONSTRAINT PRIMARY KEY("saleid, vendorid")
PARTITION BY HASH(n) KEY ("locationid");
```

When you shard a large table, you select a *shard key* from the table. This key exists in every shard.

Choosing a shard key plays an important role in the number of shards and the size of any single shard.

Best practices for choosing a shard key

Here is a full `CREATE TABLE` statement.

```

CREATE TABLE "sales_fact"
  ("saleid" int,
   "locationid" int,
   "vendorid" int,
   "quantity" int,
   "sale_amount" double,
   "fruitid" int,
   CONSTRAINT
   PRIMARY KEY("saleid", "vendorid"))
PARTITION BY HASH(96)
KEY ("saleid");

```

The shard key is a subset of the primary key. However, that is not the only guideline to follow when choosing a shard key.

- 1. If the table has a primary key, the shard key must be a subset of the primary key.**

If the shard key is **not** a subset of the primary key, and the shard key changes, data with the same primary key may reside in different nodes. This impacts ThoughtSpot's performance, and may result in incorrect query results.

You should not use a shard key that is not a subset of the primary key. If you use a shard key that is not a subset of the primary key, it is possible to get two versions of a record if the shard key for a record changes, but the primary key does not. In the absence of a unique shard key, the system creates a secondary record rather than doing a SQL MERGE (`upsert`). These two versions of a record may result in incorrect results when you search your data in ThoughtSpot.

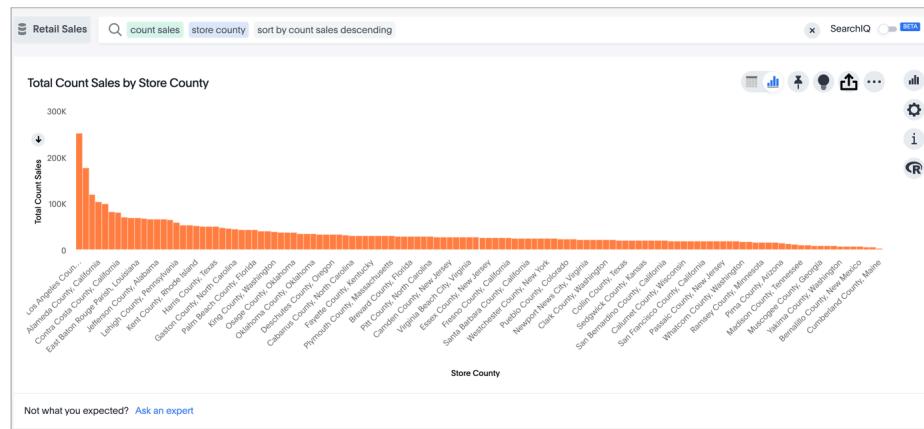
If you try to use a shard key that is not a subset of the primary key, your `CREATE TABLE` command returns an error.

- 2. Choose a shard key that distributes data well across keys.**

For example, suppose the table you want to shard has a primary key made up of `saleid`, `custid`, and `locationid`. The table has 10K sales, 400 locations, and 2000 customers. If 5K sales are in just two locations, you should not use `locationid` as your shard key. If you use `locationid` as your shard key, you have data in fewer shards, which impacts performance. Instead, you should use `custid` and `locationid`.

As a more concrete example, suppose you want to shard a table of retail data. Many retailers have an increase in sales around the winter holidays. You should not use `date` as your shard key, because you may have five or ten times your usual number of daily transactions during the month of December. Using `date` as your shard key would result in data skew, and would impact performance.

Here is an example of data skew, where `Los Angeles` has many more transactions than the average, so you should not use `store county` as your sharding key.



You may also have to clean up your data and any null values before sharding. For example, your retail data may have a `customer` column. One of the values for `customer` may be `unknown`. A value like `unknown` would exist in many more transactions than a single customer name. A value like `unknown`, or any null values, result in data skew, and impact performance.

3. Choose a shard key that results in a wide variety of keys.

For example, suppose the table you want to shard has a primary key made up of `saleid`, `productid`, and `locationid`. The table has 10K sales, 40 locations, and 200 products. Even if the sales are evenly distributed across locations, you should not use `locationid` in your shard key, because there are only 40 possible keys. Instead, use `saleid` and `productid` for more variety.

4. If you plan to join two or more tables that are both sharded, both tables must use the same shard key.

This guideline ensures better join performance. For example, if you have two tables and the primary keys are:

```
PRIMARY KEY("saleid, vendorid") on A  
PRIMARY KEY("saleid, customerid") on B
```

Use `saleid` as the shard key when you shard both tables.

5. **If your primary key includes several columns, use all appropriate columns in the shard key.**

Your primary key may include several columns. For example, suppose the table you want to shard has a primary key made up of `saleid`, `custid`, and `locationid`, as in the example in guideline three. The table has 10K sales, 40 locations, and 200 products. Based on the best practice outlined in guideline three (**choose a shard key that results in a wide variety of keys**), you should not use `locationid` in your shard key. Both `saleid` and `custid` are good shard keys, based on the four best practices mentioned above. Instead of picking one column to use as your shard key, use both `saleid` and `custid`.

You can always use your primary key as a shard key. If you have trouble picking another shard key based on the above requirements and best practices, use your primary key.

Sharded dimension tables

In a typical schema, you'd have a sharded fact table, with foreign keys to small dimension tables. ThoughtSpot replicates these small dimension tables in their entirety and distributes them on every node. If your dimension table has more than 40 million rows, however, you may want to co-shard it with related fact tables. Consult with your ThoughtSpot contact before co-sharding.

If you have a large dimension table, replicating it and distributing it can impact the performance of your ThoughtSpot system. In this case, you want to shard the dimension tables *and* the fact table. Note that you can co-shard multiple fact tables and one or more dimension tables on the same shard key. ThoughtSpot can handle chasm traps.

When sharding both a fact table and its dimension table(s), (known as co-sharding) keep in mind the guidance for creating a shard key. Only shard dimension tables if the dimension table has more than 40 million rows, and the join between the fact and dimension tables uses the same columns. Specifically, the tables must:

- be related by a primary key and foreign key
- be sharded on the same primary key/foreign key
- have the same number of shards

If these requirements are met, ThoughtSpot automatically co-shards the tables for you. Co-sharded tables are always joined on the shard key. Data skew can develop if a very large proportion of the rows have the same value for the shard key. For example, you may have an `unknown` value for a `customer` column. Many of the rows of a fact table may include this value, resulting in data skew. Refer to [sharding best practices \[See page 206\]](#) to learn how to check for data skew.

You can view your `row count skew` from the ThoughtSpot application. Go to **admin**, then **System health**, then **data**. Choose the table you would like to view, and scroll to `row count skew`. Use this number to calculate your row count skew ratio: $\text{row count skew} / (\text{total row count} / \text{number of partitions})$. A row count skew ratio higher than 1 may require changes to your data modeling.

This example shows the `CREATE TABLE` statements that meet the criteria for sharding both a fact table and its dimension table:

```

TQL> CREATE TABLE products_dim (
    "id" int,
    "prod_name" varchar(30),
    "prod_desc" varchar(100),
    PRIMARY KEY ("id")
)
PARTITION BY HASH (96) KEY ("id")
;

TQL> CREATE TABLE retail_fact (
    "trans_id" int,
    "product_id" int,
    "amount" double,
    FOREIGN KEY ("product_id") REFERENCES products_dim ("id")
)
PARTITION BY HASH (96) KEY ("product_id")
;

```

Joining two sharded fact tables

You can also join two sharded fact tables with different shard keys, but it is not recommended. This is known as *non co-sharded* tables. It may take a while to join two tables sharded on different keys, since ThoughtSpot has to redistribute your data. Therefore, ThoughtSpot recommends that you use a common shard key for two fact tables.

You are not limited by the column connection or relationship type.

Sharding best practices

There are several best practices related to sharding.

1. Shard your tables **before** loading data.

Your data loads faster if you have already sharded the tables. Use the `CREATE TABLE` command to specify how you want your tables sharded, but do not load any data. After you shard the tables, your data loads faster.

2. You may need to re-evaluate your sharding over time, as your data evolves. Take a look at how your sharding impacts performance after you change your data significantly. Data also changes naturally over time, so you should re-evaluate sharding at a regular cadence.

To evaluate your sharding strategy, run the following script. It checks for over- or under-sharded tables on your cluster.

- a. Log into your cluster on the command line.

```
$ ssh admin@<cluster-IP>
```

- b. Run the following script to check for over- or under-sharded tables.

```
$ /usr/local/scaligent/release/bin/sharding_di  
agnostics.sh
```

- c. Adjust your sharding strategy appropriately. See [Change sharding on a table \[See page 231\]](#).

Note that resharding automatically loads data into a new incarnation of the table you sharded. You do not need to reload the table's data.

3. Check your `row count skew` ratio when you re-evaluate sharding.

You can view your `row count skew` from the ThoughtSpot application. Go to **admin**, then **System health**, then **data**. Choose the table you would like to view, and scroll to `row count skew`. Use this number to calculate your row count skew ratio: `row count skew / (total row count / number of partitions)`. A row count skew ratio higher than 1 may require changes to your data modeling.

Chasm traps

Summary: A chasm trap occurs when two many-to-one joins converge on a single table.

In a complex schema, you may have a fact table with no relationship to another fact table, except that each contains a foreign key to a shared dimension table. This is known as a chasm trap, and ThoughtSpot can handle it!

Understand how chasm traps occur

A fact table, just as it sounds, stores facts about your business. If you are selling apples, the sales fact table has facts about these apples.

SaleID	AppleTypeID	StoreID	Units Sold
4	55	2	12
8	34	33	3
10	09	09	1

Dimension tables describe the attributes that are interesting to analyze. For example, the apple table might look like this.

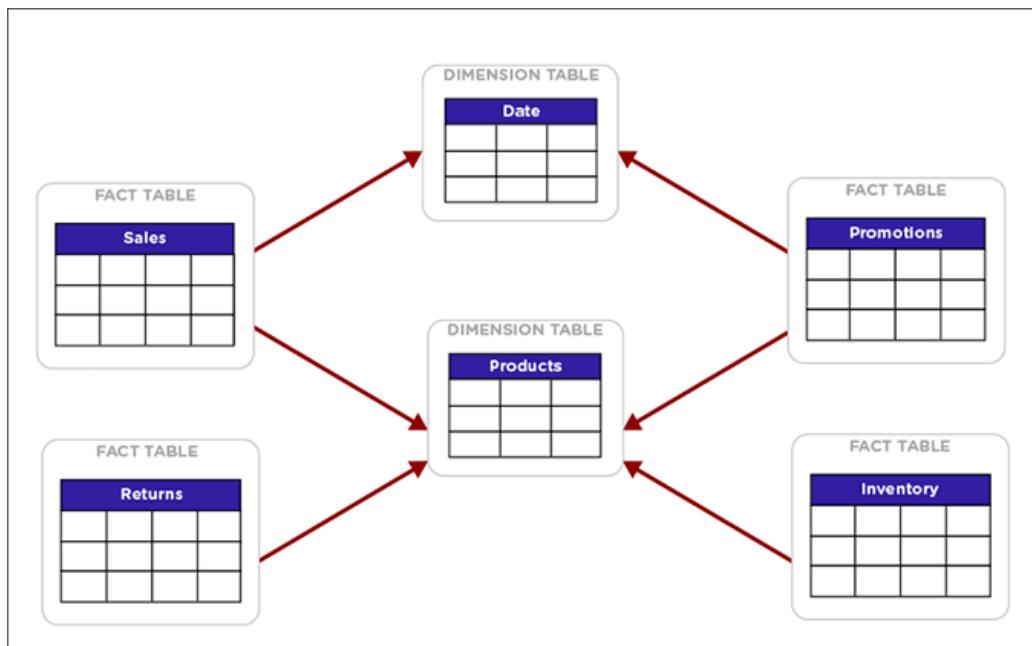
AppleTypeID	Color	Name	Use
55	Red	Red Delicious	Snack
34	Green	Granny Smith	Cooking
09	Yellow	Golden	Snack

As you can imagine, in a business you might have several fact tables that access dimension tables. So, an apple business may record waste as well as sales.

TimeID	AppleTypeID	StoreID	Units Wasted
4	55	2	2
8	34	33	43
10	09	09	11

Both the sales and waste tables are facts that reference the apple dimension table.

A chasm trap in a data schema can introduce problems of over counting if you join the two fact tables through their shared dimension table. This diagram shows a typical complex schema with several tables that are related over a chasm trap:



Examples of use cases where a chasm trap could occur when attribution analysis compare campaign data with purchase data, where all they have in common is that both contain a customer identifier that is a foreign key to a customer dimension table. Chasm traps also occur, for example, in cost of sales analysis when wholesale orders data is only related to the retail sales data through a shared products dimension table.

In many databases, joining tables across a chasm trap creates a *Cartesian product* or *cross join*. That is each row from the first fact table is joined to each row from the second table. A Cartesian product causes over counting when computing counts and aggregates. ThoughtSpot protects you from this kind of over counting.

There are still just a few things to look out for when using a schema that contains chasm traps:

- The tables should be joined to the dimension table by an equi-join (a primary key/foreign key relationship). They cannot be joined using a range of values.
- Review the column setting called [Attribution Dimension \[See page 444\]](#). You may need to change this setting if some of the columns in the shared dimension table should not be used for attribution when combining fact tables.
- Tables that will be joined across a chasm trap do not have to be co-sharded. They will be joined appropriately automatically in the most efficient way.

Chasm trap limitations

Join information in **What am I Looking At?** does not appear for searches on a worksheet containing a chasm trap or on base tables that are related over a chasm trap.

Overview of schema building

Summary: Before you can load data into ThoughtSpot, you must build a database schema to receive it.

You can build a schema by writing a SQL script that creates the objects in your schema. ThoughtSpot provides the ThoughtSpot SQL Command Line (TQL) for creating, viewing, and managing a schema using SQL. Your script can use any statements that are supported in ThoughtSpot SQL Command Line (TQL). The TQL syntax is similar to the SQL used in other relational databases, but with some important differences.

You can use DDL (data definition language) to create the schema into which you can load the data. We recommend placing all your DDL statements into a single script for creating the schema.

Upload the script through the browser

You can upload an your SQL script directly through the browser in the ThoughtSpot application. You can edit the script or add to it right within the browser, too. The steps to build a schema through the browser are:

1. [Write a SQL script to create the schema \[See page 215\]](#)
2. [Import a schema \(use the SQL editor\) \[See page 222\]](#)

Use TQL on the SQL command line

You can choose to run your SQL script within the Linux shell instead. You can run TQL in interactive command line mode, or you can write a script and use TQL to run it. The SQL syntax in ThoughtSpot is called TQL for ThoughtSpot SQL. The ThoughtSpot SQL Command Line (TQL) runs in an interactive mode. To invoke TQL Log in to the Linux shell using SSH and type `tql`. At the prompt, type `h` or `help` to see a list of supported commands.

The steps to build a schema using TQL include:

1. [Connect to the database with the ThoughtSpot SQL Command Line \(TQL\) \[See page 213\].](#)
2. [Write a SQL script to create the schema \[See page 215\].](#)

3. Type your SQL commands on the command line, terminating each command with a semicolon (;).

Commands can span multiple lines. ThoughtSpot supports a limited number of SQL commands, plus some custom SQL extensions. For example, you can specify the number of shards and the distribution key as part of the CREATE TABLE syntax. A full list of supported SQL in TQL is available in the [TQL reference \[See page 0\]](#).

Where to go next

- [Connect with TCL and create a schema \[See page 213\]](#)

Having examined the structure of the data to be loaded and become familiar with the ThoughtSpot SQL Command Line (TQL), you are now ready to create the schema.

- [Write a SQL script to create the schema \[See page 215\]](#)

Using a SQL script to create your schema is a recommended best practice. This makes it easier to adjust the schema definitions and recreate the schema quickly, if needed.

- [Schema creation examples \[See page 217\]](#)

These examples demonstrate the steps involved in creating a schema using the ThoughtSpot SQL Command Line (TQL). After the schema is created, you can load data into it with ThoughtSpot Loader.

- [Upload and run a SQL script\) \[See page 222\]](#)

You can run a SQL script to create your database schema through the browser, without having to log in to the shell on the ThoughtSpot instance. You can edit the script and run it directly in the browser to create the schema.

Connect with TQL and create a schema

Summary: Learn about TQL, ThoughtSpot's SQL command line.

To perform administrative tasks directly in the database, you will use the ThoughtSpot SQL Command Line (TQL). TQL supports many, but not all, common SQL commands.

Connect with TQL

Before connecting with TQL, you need:

- Access to your ThoughtSpot instance Linux shell from a client machine.
- The administrator OS login.

To connect to TQL:

1. Log in to the Linux shell using SSH.
2. Invoke TQL:

```
$ tql  
TQL>
```

3. Enter your SQL command, followed by a semicolon (;).

Enter a SQL script

Having examined the structure of the data to be loaded and become familiar with the ThoughtSpot SQL Command Line (TQL), you are now ready to create the schema.

This method is a good way to get familiar with TQL and how to create database objects, but when creating a schema in a production system, you will most likely [Write a SQL script to create the schema \[See page 215\]](#).

To create the schema directly in TQL:

1. Connect to the database with the ThoughtSpot SQL Command Line (TQL) [See page 213].
2. If the database you will be using does not exist, create it now:

```
TQL> CREATE DATABASE my_database;
```

3. Connect to the database:

```
TQL> USE my_database;
```

4. If you wish to use a schema other than the default one, create it now:

```
TQL> CREATE SCHEMA my_schema;
```

5. Issue a `CREATE TABLE` command for each table you will create, using the information in [Plan the schema](#) [See page 184].

Tip: Foreign key declaration within a `CREATE TABLE` will show the table created even if there are problems with the foreign key. Therefore, it is good practice to also issue a separate `ALTER TABLE ADD CONSTRAINT FOREIGN KEY` command.

Creating a schema using SQL

Summary: Using a SQL script to create your schema is a recommended best practice. This makes it easier to adjust the schema definitions and recreate the schema quickly, if needed.

The script for creating a schema is a text file that contains all the SQL commands to create your schema.

Use the comment tags /* and */ to document your script.

Enclose all object names (schema, table, and column) in double quotes, and any column values in single quotes. Use double quotes for object names that are also reserved words in SQL, or that contain special characters. Special characters are all characters other than letters, numbers, or the underscore (_). If you get the error message “Error parsing SQL. Check SQL input.”, check your script for object names without double quotes.

If you are working in a schema other than the default schema, object names must be fully qualified:

```
"<schema_name>"."<object_name>" .
```

If your schema includes constraints to define relationships between tables, such as primary key and foreign key definitions, or the RELATIONSHIP syntax, we recommend that you first create all the tables, and then add the relationships between the tables using the ADD CONSTRAINT syntax. This approach makes it easier to troubleshoot the script and make subsequent changes.

1. Open a new file in a text editor.
2. Type in the command to create the database, if it does not already exist:

```
CREATE database <db_name>;
```

3. Type in the command to specify the database to use:

```
USE database <db_name>;
```

4. Type in the command to create the schema, if you don't want to use the default schema:
5. Type in each of the CREATE TABLE statements, with its column definitions, primary key

constraints, and sharding specification (if any).

6. At the end of your script, optionally type in the `ALTER TABLE` statements to add foreign keys to use in joining the tables.
7. Save the file.
8. Run the script using one of these methods:
 - [Import a schema \(use the SQL editor\) \[See page 222\]](#).
 - [Log in to the shell \[See page 9\]](#), copy your script to your ThoughtSpot instance using `scp`, and pipe it to TQL:

```
$ cat create-schema.sql | tql
```

Schema creation examples

Summary: Use the TQL and the ThoughtSpot Loader.

These examples demonstrate the steps involved in creating a schema using the ThoughtSpot SQL Command Line (TQL). After the schema is created, you can load data into it with ThoughtSpot Loader.

Simple schema creation example

The example creates a database (`tpch`) with two tables (`customer`, `transaction`). The example does not create a schema explicitly. So it will use the default schema (`falcon_default_schema`).

In this example:

- The table `customer` has a primary key called `customer_id`. The table `customer_transactions` has a primary key called `transaction_id`.
- The `customer` table is unsharded.
- The `customer_transactions` table is sharded into 96 shards using the `transaction_id` column.
- Both tables have referential integrity on `customer_id`.

```
$tql

TQL> CREATE DATABASE tpch;

TQL> USE tpch;

TQL> CREATE TABLE customer (
    name  VARCHAR(100),
    address VARCHAR(255),
    zipcode INT,
    customer_id INT,
    CONSTRAINT PRIMARY KEY (customer_id)
);

TQL> CREATE TABLE customer_transactions (
    transaction_id INT,
    customer_id INT,
    amount DOUBLE,
    transaction_date DATETIME,
    CONSTRAINT PRIMARY KEY (transaction_id),
    CONSTRAINT FOREIGN KEY (customer_id) REFERENCES
    customer(customer_id)
) PARTITION BY HASH (96) KEY (transaction_id);
```

More complex schema creation example

The example uses a custom schema called sample_schema to hold the tables. Because of this, every table reference has to be schema qualified.

```
$ tql

TQL> CREATE DATABASE "sample_db";

TQL> USE "sample_db";

TQL> CREATE SCHEMA "sample_schema";

TQL> CREATE TABLE "sample_schema"."customer" (
    "c_custkey" BIGINT,
    "c_name" VARCHAR(255),
    "c_address" VARCHAR(255),
    "c_city" VARCHAR(255),
    "c_nation" VARCHAR(255),
    "c_region" VARCHAR(255),
    "c_phone" VARCHAR(255),
    CONSTRAINT PRIMARY KEY ("c_custkey")
);

TQL> CREATE TABLE "sample_schema"."supplier" (
    "s_suppkey" BIGINT,
    "s_name" VARCHAR(255),
    "s_address" VARCHAR(255),
    "s_city" VARCHAR(255),
    "s_nation" VARCHAR(255),
    "s_region" VARCHAR(255),
    "s_phone" VARCHAR(255),
```

```
CONSTRAINT PRIMARY KEY ("s_suppkey")  
);  
  
TQL> CREATE TABLE "sample_schema"."lineorder" (  
    "lo_orderkey" BIGINT,  
    "lo_linenumber" BIGINT,  
    "lo_custkey" BIGINT,  
    "lo_partkey" BIGINT,  
    "lo_suppkey" BIGINT,  
    "lo_orderdate" DATE,  
    "lo_orderpriority" VARCHAR(255),  
    "lo_shipppriority" VARCHAR(255),  
    "lo_quantify" BIGINT,  
    "lo_extendprice" BIGINT,  
    "lo_ordtotalprice" BIGINT,  
    "lo_discount" BIGINT,  
    "lo_commitdate" DATE,  
    CONSTRAINT PRIMARY KEY ("lo_orderkey","lo_linenumber"),  
    CONSTRAINT FOREIGN KEY ("lo_custkey") REFERENCES "sample_sche  
ma"."customer" ("c_custkey"),  
    CONSTRAINT FOREIGN KEY ("lo_suppkey") REFERENCES "sample_sche  
ma"."supplier" ("s_suppkey")  
) PARTITION BY HASH (96) KEY (lo_orderkey);
```

Upload and run a SQL script

Summary: You can import a schema through the Web browser, allowing you to run your SQL script without SSH access for the cluster.

You can run a SQL script to create your database schema and relationships or joins through the browser, without having to log into the shell on the ThoughtSpot instance. You can edit the script and run it directly in the browser to create the schema. Note that you must have the **can manage data** permission and **admin** privileges to upload schemas. Use this feature for the following actions:

- Create the SQL script ahead of time [See page 215], and [use the browser to run it \[See page 222\]](#).
- Use the **Write TQL** interface [See page 225] to type your SQL directly into the browser.
- Use the **Write TQL** interface as an interactive SQL editor, to make changes to an existing schema.

Run or edit an existing script

To run or edit a SQL script that you already wrote to create your database schema and relationships through the browser, follow these steps.

1. Log in to ThoughtSpot from a browser.
2. Click **Data**, on the top navigation bar.
3. Click the ellipsis icon  , and select **Upload schema**.

Upload and run a SQL script

The screenshot shows the ThoughtSpot interface under the 'DATA' tab, specifically the 'Data Objects' section. A context menu is open over the 'NewRetail_Product_Dimension' row, with the 'Upload schema' option highlighted.

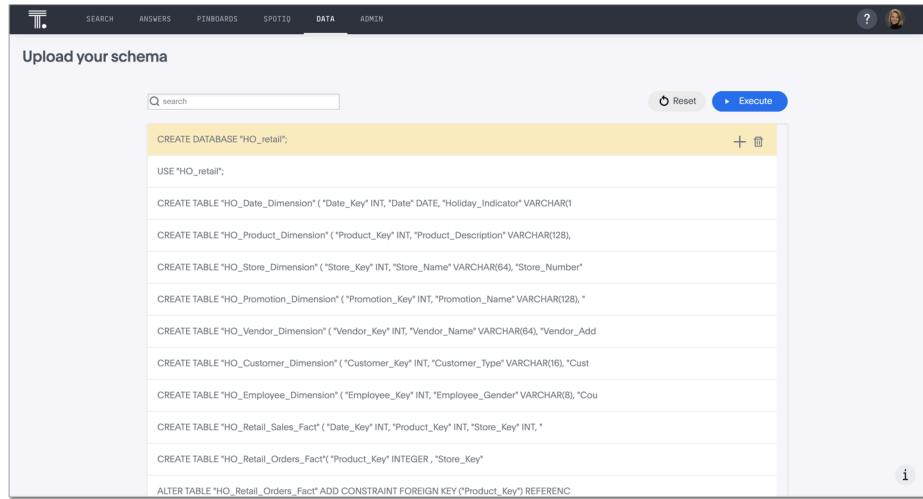
Name	Source	Stickers	Materialize Status	Modified	User
NewRetail_Product_Dimension				a year ago	
NewRetail_Baskets_Fact				a year ago	Administrator Super-User
NewRetail_Sales_Fact				a year ago	Administrator Super-User
NewRetail_Product_Cost_Dimension				a year ago	Administrator Super-User
NewRetail_Store_Dimension				a year ago	Administrator Super-User
NewRetail_Product_Dimension				a year ago	Administrator Super-User
NewRetail_Customer_Dimension				a year ago	Administrator Super-User

4. Drag and drop your SQL script into the browser, or choose **Browse Your Files** to locate it.

You cannot copy and paste an existing SQL script into the SQL editor.

The screenshot shows the ThoughtSpot SQL editor interface. The title bar says 'Upload your schema'. At the top right are buttons for 'Write TSQL', 'Reset', and 'Execute'. Below is a large dashed rectangular area for file upload, with a 'Browse your files...' button and a note 'Maximum upload file size: 50MB'.

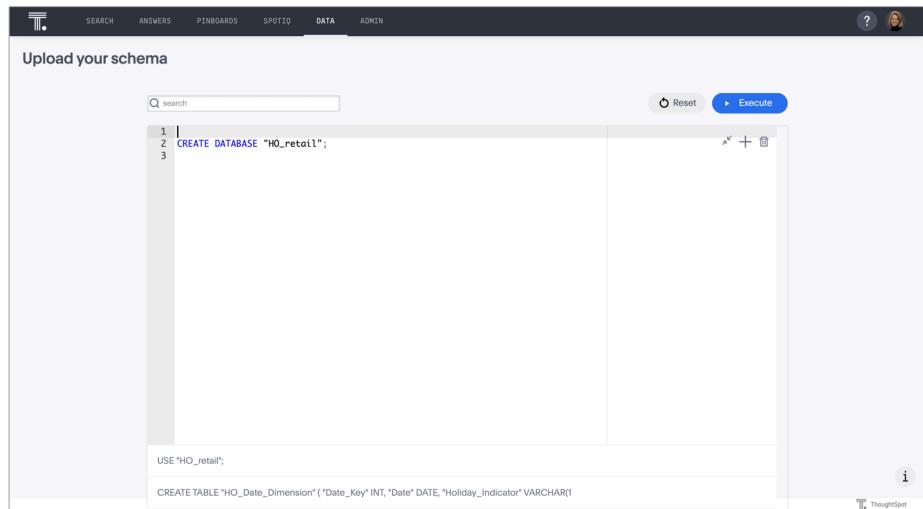
5. After you upload your SQL script, you enter the SQL editor. Use it to view your script and make any changes.



The screenshot shows the ThoughtSpot interface with the "DATA" tab selected. The main area is titled "Upload your schema". A search bar is at the top left, followed by a "Reset" button and an "Execute" button. The code editor contains a large block of SQL code:

```
CREATE DATABASE "HO_retail";
USE "HO_retail";
CREATE TABLE "HO_Date_Dimension" ("Date_Key" INT, "Date" DATE, "Holiday_Indicator" VARCHAR(1);
CREATE TABLE "HO_Product_Dimension" ("Product_Key" INT, "Product_Description" VARCHAR(128),
CREATE TABLE "HO_Store_Dimension" ("Store_Key" INT, "Store_Name" VARCHAR(64), "Store_Number"
CREATE TABLE "HO_Promotion_Dimension" ("Promotion_Key" INT, "Promotion_Name" VARCHAR(128),
CREATE TABLE "HO_Vendor_Dimension" ("Vendor_Key" INT, "Vendor_Name" VARCHAR(64), "Vendor_Add
CREATE TABLE "HO_Customer_Dimension" ("Customer_Key" INT, "Customer_Type" VARCHAR(16), "Cust
CREATE TABLE "HO_Employee_Dimension" ("Employee_Key" INT, "Employee_Gender" VARCHAR(8), "Cou
CREATE TABLE "HO_Retail_Sales_Fact" ("Date_Key" INT, "Product_Key" INT, "Store_Key" INT, "Sales
CREATE TABLE "HO_Retail_Orders_Fact" ("Product_Key" INTEGER, "Store_Key" INT, "Order_Status" V
ALTER TABLE "HO_Retail_Orders_Fact" ADD CONSTRAINT FOREIGN KEY ("Product_Key") REFERENCES "HO_Promotion_Dimension" ("Promotion_Key");
```

6. To edit the script, click on the line you would like to change, and type your changes into the interface.



The screenshot shows the ThoughtSpot interface with the "DATA" tab selected. The main area is titled "Upload your schema". A search bar is at the top left, followed by a "Reset" button and an "Execute" button. The code editor has the first line of the script highlighted in blue:

```
1 | CREATE DATABASE "HO_retail";
2 |
3 |
```

7. When ready, run your script by clicking the **Execute** button.

The screenshot shows the ThoughtSpot Data interface. At the top, there's a navigation bar with links for SEARCH, ANSWERS, PINBOARDS, SPOTIO, DATA, and ADMIN. Below the navigation bar, a search bar is followed by a 'Reset' button and an 'Execute' button, which is highlighted with a mouse cursor. The main area contains a code editor with several lines of SQL script. The script creates a database named 'HO_retail', uses it, and then creates various dimension tables (Date, Product, Store, Promotion, Vendor, Customer, Employee) and fact tables (Sales_Fact, Orders_Fact). The code editor has a search bar at the top and a toolbar with icons for copy, paste, and execute.

```

CREATE DATABASE "HO_retail";
USE "HO_retail";

CREATE TABLE "HO_Date_Dimension" ("Date_Key" INT, "Date" DATE, "Holiday_Indicator" VARCHAR(1);

CREATE TABLE "HO_Product_Dimension" ("Product_Key" INT, "Product_Description" VARCHAR(128);

CREATE TABLE "HO_Store_Dimension" ("Store_Key" INT, "Store_Name" VARCHAR(64), "Store_Number"

CREATE TABLE "HO_Promotion_Dimension" ("Promotion_Key" INT, "Promotion_Name" VARCHAR(128);

CREATE TABLE "HO_Vendor_Dimension" ("Vendor_Key" INT, "Vendor_Name" VARCHAR(64), "Vendor_Add

CREATE TABLE "HO_Customer_Dimension" ("Customer_Key" INT, "Customer_Type" VARCHAR(16), "Cust

CREATE TABLE "HO_Employee_Dimension" ("Employee_Key" INT, "Employee_Gender" VARCHAR(8), "Cust

CREATE TABLE "HO_Retail_Sales_Fact" ("Date_Key" INT, "Product_Key" INT, "Store_Key" INT, "Q

CREATE TABLE "HO_Retail_Orders_Fact" ("Product_Key" INTEGER, "Store_Key"

```

- If there are any errors, correct them and run the script again.

Write a new SQL script

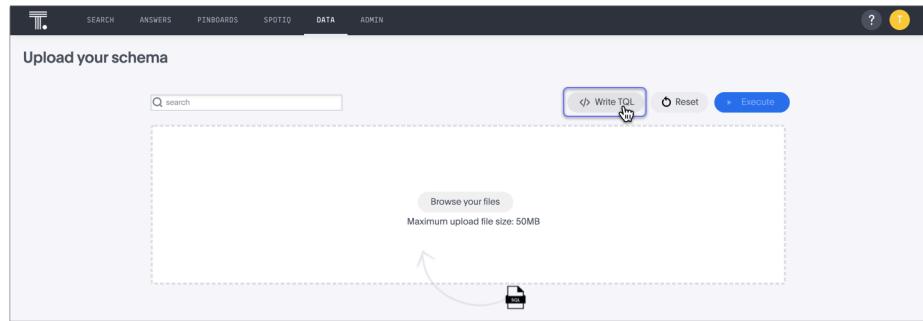
To write a new SQL script to create your database schema and relationships through the browser, follow these steps.

- Log in to ThoughtSpot from a browser.
- Click **Data**, on the top navigation bar.
- Click the ellipsis icon , and select **Upload schema**.

The screenshot shows the ThoughtSpot Data Objects page. The top navigation bar includes links for SEARCH, ANSWERS, PINBOARDS, SPOTIO, DATA, and ADMIN. Below the navigation bar, there are tabs for Data Objects, Connections, Settings, Usage, and DataFlow. Under the Data Objects tab, there are filters for All, Yours, All types, Worksheets, Tables, and Views. On the right side, there's a sidebar with a 'Stickers' section and a 'Create worksheet' button. The main area displays a list of data objects, each with a name, source, stickers, materialize status, modified date, and owner. A context menu is open over the first item in the list, showing options like Create worksheet, Create worksheet from file, Upload data, and Upload schema. The 'Upload schema' option is highlighted with a blue selection bar and a mouse cursor. The data objects listed are NewRetail_Product_Dict_Dimension, NewRetail_Baskets_Fact, NewRetail_Sales_Fact, NewRetail_Product_Cost_Dimension, NewRetail_Store_Dimension, NewRetail_Product_Dimension, and NewRetail_Customer_Dimension, all created a year ago by a user with the initials 'A'.

Name	Source	Stickers	Materialize Status	Modified	Owner
NewRetail_Product_Dict_Dimension				a year ago	
NewRetail_Baskets_Fact				a year ago	(A) Administrator Super-User
NewRetail_Sales_Fact				a year ago	(A) Administrator Super-User
NewRetail_Product_Cost_Dimension				a year ago	(A) Administrator Super-User
NewRetail_Store_Dimension				a year ago	(A) Administrator Super-User
NewRetail_Product_Dimension				a year ago	(A) Administrator Super-User
NewRetail_Customer_Dimension				a year ago	(A) Administrator Super-User

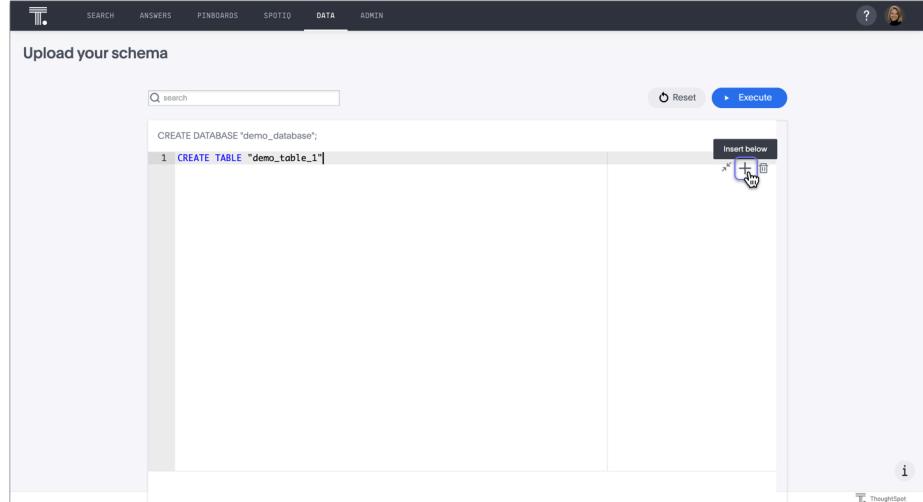
4. Select the **Write TQL** button in the top right corner of the screen.



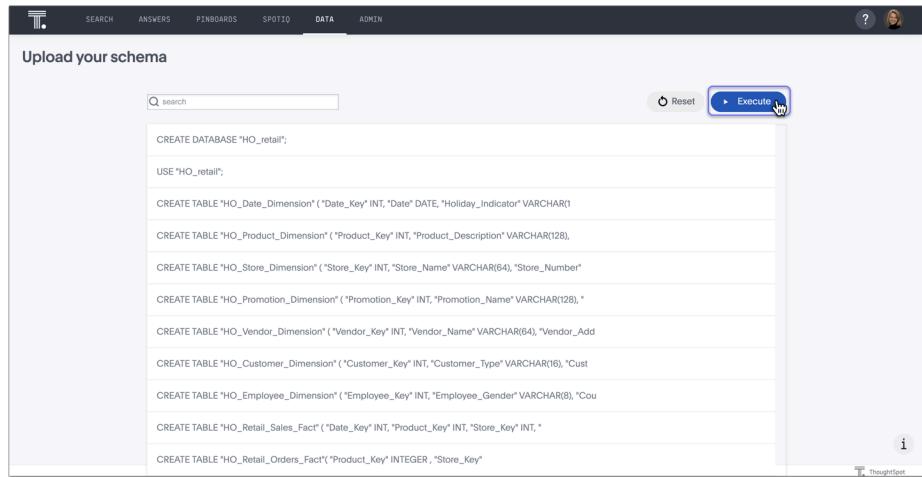
5. After you select **Write TQL**, you enter the SQL editor. You can write your script for schema and relationship creation in this editor.

Note: You cannot copy and paste an existing SQL script into the SQL editor.

When you finish writing a line, click the + icon to insert a new line below.



6. When ready, run your script by clicking the **Execute** button.



The screenshot shows the ThoughtSpot interface with the "DATA" tab selected. A search bar at the top has "search" typed into it. Below the search bar, there is a large text area containing a SQL script. The script creates a database named "HO_retail" and defines several tables: HO_Date_Dimension, HO_Product_Dimension, HO_Store_Dimension, HO_Promotion_Dimension, HO_Vendor_Dimension, HO_Customer_Dimension, HO_Employee_Dimension, HO_Retail_Sales_Fact, and HO_Retail_Orders_Fact. The "Execute" button is highlighted with a blue border and a cursor arrow pointing to it. The "Reset" button is also visible.

```
CREATE DATABASE "HO_retail";
USE "HO_retail";
CREATE TABLE "HO_Date_Dimension" ("Date_Key" INT, "Date" DATE, "Holiday_Indicator" VARCHAR(1));
CREATE TABLE "HO_Product_Dimension" ("Product_Key" INT, "Product_Description" VARCHAR(128));
CREATE TABLE "HO_Store_Dimension" ("Store_Key" INT, "Store_Name" VARCHAR(64), "Store_Number" INT);
CREATE TABLE "HO_Promotion_Dimension" ("Promotion_Key" INT, "Promotion_Name" VARCHAR(128), "Promotion_StartDate" DATE, "Promotion_EndDate" DATE);
CREATE TABLE "HO_Vendor_Dimension" ("Vendor_Key" INT, "Vendor_Name" VARCHAR(64), "Vendor_Address" TEXT, "Vendor_City" TEXT, "Vendor_State" TEXT, "Vendor_PostalCode" TEXT, "Vendor_Country" TEXT);
CREATE TABLE "HO_Customer_Dimension" ("Customer_Key" INT, "Customer_Type" VARCHAR(16), "Customer_Gender" VARCHAR(8), "Customer_AgeGroup" VARCHAR(16));
CREATE TABLE "HO_Employee_Dimension" ("Employee_Key" INT, "Employee_Gender" VARCHAR(8), "Employee_HireDate" DATE, "Employee_MaritalStatus" VARCHAR(16));
CREATE TABLE "HO_Retail_Sales_Fact" ("Date_Key" INT, "Product_Key" INT, "Store_Key" INT, "Sales_Qty" INT, "Sales_Amt" DECIMAL(10, 2));
CREATE TABLE "HO_Retail_Orders_Fact" ("Product_Key" INTEGER, "Store_Key" INT, "Order_Qty" INT, "Order_Amt" DECIMAL(10, 2));
```

7. If there are any errors, correct them and run the script again.

How to change a schema

Summary: After creating the schema and loading data, you may have to change your set up. You can alter the schema by changing the primary key, relationships between tables, and table sharding.

Making changes to a schema after data has been loaded and users have created worksheets or pinboards on the tables requires care, so that you don't lose the relationship between the objects created in ThoughtSpot and the underlying tables. If you follow the procedures here, your tables will retain their relationships to the objects created on top of them.

Tip: Always take a snapshot of your database before making any schema changes. This snapshot allows you to revert back to the prior state if you make an error, or something doesn't work as you expected after the schema change.

Change the primary key for a table

Use this procedure to change the primary key for a table. But use it with caution, particularly if you are changing to a primary key for which values are not unique.

You can change the primary key of a table without having to `TRUNCATE` it first and reload the data. However, changing the primary key could result in data deletion. This is because of the upsert behavior which is applied when multiple rows have the same primary key. This is very important to understand ahead of time, if you are considering changing to a primary key for which values are not unique.

To change the primary key, first remove any existing primary key, and then define the new key. If you no longer want the table to have a primary key, just remove the existing key. Any dependent objects, like Pinboards or Worksheets, remain intact.

To change the primary key of a table:

1. [Create a manual snapshot \[See page 385\]](#).
2. [Connect to the database with the ThoughtSpot SQL Command Line \(TQL\) \[See page 213\]](#).
3. Drop the existing primary key (if any), by issuing a command like this example:

```
TQL> ALTER TABLE "cart"
      DROP CONSTRAINT
      PRIMARY KEY;
```

Dropping a primary key can impact existing Worksheets, Answers, and Pinboards. The system warns you if dropping a primary key impacts other objects. [//]: # (To continue, use the `--allow_unsafe` flag.)

4. Add a new primary key, if desired:

```
TQL> ALTER TABLE "cart"
      ADD CONSTRAINT
      PRIMARY KEY ("owner_id");
```

5. Test that any dependent objects (Pinboards, Worksheets, and so on) are still working correctly.
6. Delete the snapshot you created earlier:

```
tscli snapshot delete <name>
```

Change a relationship between tables

Use this procedure to remove a relationship between tables or define a new one. This operation works for both kinds of relationships: foreign key or generic relationship.

To change a relationship between two tables, first remove any existing relationship, and then define the new relationship (if any). You do not have to truncate the tables to do this operation. Any dependent objects (pinboards or worksheets) will remain intact.

To change the relationship between tables:

1. [Create a manual snapshot \[See page 385\]](#).
2. [Connect to the database with the ThoughtSpot SQL Command Line \(TQL\) \[See page 213\]](#).
3. Issue the command to drop the existing relationship

Before dropping a relationship TQL checks for and then warns of any dependent objects. [//]:

(To continue with the drop any way, use the `--allow_unsafe` flag.)

The following examples illustrate several different types of drop operations.

Drop a foreign key by name, if it was given a name when it was defined:

```
TQL> ALTER TABLE  
    "sales_fact"  
    DROP CONSTRAINT  
    "FK_P0_number";
```

Drop a relationship by name, if it was given a name when it was defined:

```
TQL> ALTER TABLE "fruit_dim"  
    DROP RELATIONSHIP "REL_dates";
```

Drop the foreign key relationship explicitly, if it doesn't have a name, by referencing the two tables that are joined. This drops all foreign keys between the two tables:

```
TQL> ALTER TABLE "shipments"  
    DROP CONSTRAINT  
    FOREIGN KEY "orders";
```

Drop all generic relationships between two tables:

```
TQL> ALTER TABLE "wholesale_buys"  
    DROP RELATIONSHIP  
    WITH "retail_sales";
```

4. Define a new relationship, if you want to, using `ALTER TABLE...ADD CONSTRAINT...`
5. Test that any dependent objects (pinboards, worksheets, etc.) perform correctly.
6. Delete the snapshot you created earlier using the command:

```
tscli snapshot delete <name>
```

Change sharding on a table

You can change the sharding on a table or remove it altogether (creating a replicated table) using this procedure. This procedure preserves the data within the table.

This procedure reshards a table. This is also called redistributing or repartitioning. You can use this method to reshuffle a table without losing its data or metadata. This means that worksheets and pinboards built on top of the table will continue to work.

You can use these steps to do any of these operations:

- shard a table that was previously replicated.
- change a replicated table to a sharded table.
- change the number of shards to use for a sharded table. This is useful if the size of your cluster has increased.

To change the sharding on a table:

1. [Create a manual snapshot \[See page 385\]](#).
2. [Connect to the database with the ThoughtSpot SQL Command Line \(TQL\) \[See page 213\]](#).
3. Issue the command to change the sharding using this syntax:

```
TQL> ALTER TABLE <table>
      [SET DIMENSION | SET FACT
      [PARTITION BY HASH
      [<shards>]]
      [KEY(<column>)]]]
```

For example:

- To make a sharded table into a dimension table (replicated on every node), use:

```
ALTER TABLE "products"  
    SET DIMENSION;
```

- To make a dimension table into a sharded (fact) table or change the number of shards, use:

```
ALTER TABLE "sales"  
    SET FACT PARTITION BY HASH (96)  
    KEY ("productID");
```

Note: When you re-shard the table, the system re-shards all the existing data. You do not need to force reload.

4. Test that any dependent objects (pinboards, worksheets, etc.) are still working correctly.
5. Delete the snapshot you created earlier using the command:

```
tscli snapshot delete <name>
```

Convert column data type

Summary: You can convert the data in a column from one data type to another by issuing a TQL command.

There are some details you should be aware of when doing a data type conversion.

Data type conversion behavior

When converting from one data type to another, ThoughtSpot sets any values that it cannot convert to `NULL`. If errors occur during data type conversion, the operation terminates.

Multiple columns of a single table can be converted using a single TQL command. The behavior is transactional. So for example, you would issue the following command:

```
ALTER TABLE products
    MODIFY COLUMN product_id int,
    MODIFY COLUMN supplier VARCHAR(4);
```

Note that changing data type has implications on the primary key and sharding enforcement. For example, changing the data type of a column that is part of the sharding key would lead to a redistribution of data. Then imagine that the sharding key column contained the text values `00100`, `0100`, and `100`, which all map to same integer value. If this type of a column is changed from a `VARCHAR` to an `INT`, then it would be subject to the upsert behavior on primary keys. So, in this example, only one of the three rows would be preserved.

Be aware that data type conversion will preserve the data in the underlying database table, but there is no guarantee that any objects built on top of it (worksheets or pinboards) will be preserved. This is because you might make a data type change that makes a chart built on top of the table invalid (for example a growth chart would be invalidated if the date column it depends on were changed to a varchar column).

Supported data type conversions

In general, the data type conversions that make logical sense are supported. But there are a few nuances you should be aware of:

- When you convert from `INT` to `BOOL`, zero is converted to false, and all non-zero values are converted to true.
- When you convert from `BOOL` to `INT`, true gets converted to 1, and false gets converted to 0.
- When you convert from `DOUBLE` to `INT`, the value gets rounded.
- When you convert from `INT` to `DOUBLE`, the value gets rounded.
- When you convert from `DATETIME` to `DATE`, the date part of value is preserved and the time part is dropped.
- When you convert from `DATE` to `DATETIME`, the time gets added as `00:00:00`. The date part of the value is preserved.
- When you convert from `DATETIME` to `TIME`, the time part of the value is preserved.
- Conversion from `TIME` to `DATETIME` is not supported.

Date and time conversions

Some data type conversion require a format string. These include:

- conversion from `DATE` / `TIME` / `DATETIME`
- conversion to `DATE` / `TIME` / `DATETIME`

For these types of conversions, you can use a special syntax using `parsinghint` and the date format specifications supported in the [strftime library function](http://man7.org/linux/man-pages/man3/strftime.3.html) (<http://man7.org/linux/man-pages/man3/strftime.3.html>).

For the example, first create a table with a timestamp stored as a `VARCHAR`:

```
CREATE TABLE fruit_sales  
  (time_of_sale VARCHAR(32));  
  
INSERT INTO fruit_sales  
  VALUES ('2015-12-29 13:52:39');
```

Now, convert the column from a `VARCHAR` to `DATETIME`, using the format `%Y-%m-%d %H:%M:%S`:

```
ALTER TABLE fruit_sales  
  MODIFY COLUMN time_of_sale DATETIME  
  [parsinghint="%Y-%m-%d %H:%M:%S"]
```

Finally, convert the column back to `VARCHAR`:

```
ALTER TABLE fruit_sales  
  MODIFY COLUMN time_of_sale VARCHAR(32);
```

String to boolean conversions

String to boolean conversions have format strings, too. You can use `parsinghint` as you do for date and time conversions. You can choose among these approaches:

OPTION 1: Specify string values for both true and false. Any non-matching values get converted to null. In this example, “100” gets converted to true, and “0” gets converted to false. “-1” gets converted to null.

```
ALTER TABLE db  
  MODIFY COLUMN s bool [parsinghint="100_0"];
```

OPTION 2: Specify a string value for true. Any non-matching value gets converted to false. In this example, “100” gets converted to true, “-1” and “0” get converted to false.

```
ALTER TABLE db
    MODIFY COLUMN s bool [parsinghint="100_"];
```

Option 3: Specify a string value for false. Any non-matching value get converted to true. In this example, “-1” and “100” get converted to true, and “0” gets converted to false.

```
ALTER TABLE db
    MODIFY COLUMN s bool [parsinghint="_0"];
```

String to boolean conversions

When converting from a string to a boolean, you must specify a string for true and false. By default, a string to boolean conversion generates `true` for `true`, `false` for `false`.

```
ALTER TABLE db
    MODIFY COLUMN b varchar(32);
```

But you may override the default strings that get generated by using `parsinghint`, as in this example:

```
ALTER TABLE db
    MODIFY COLUMN b varchar(32) [parsinghint="tr_fa"];
```

Change the Data Type of a Column

When you issue the TQL command to convert a column from one data type to another, the conversion is handled automatically. However, you must ensure that any visualizations built on top of the table display correctly.

You should always take a snapshot of your database before making any schema changes. This will allow you to revert back to the prior state if you make an error, or something doesn’t work as you expected after the schema change.

When changing a data type in an existing table, be aware that answers and pinboards created on top of that table (or worksheets that include it) may change. This is because charts and aggregations depend upon the data type. So for example changing from `INTEGER` to `VARCHAR` could break charts that used the numeric data type `INTEGER` to calculate an average or a total. Because of this, use caution, and check all dependent objects before and after changing the data type, to ensure that they display as intended.

To change the data type of a column:

1. [Connect to the database with the ThoughtSpot SQL Command Line \(TQL\) \[See page 213\].](#)
2. Issue the command to change the data type using this syntax:

```
TQL> ALTER TABLE <table>
    MODIFY COLUMN <column> <new_data_type>;
```

For example:

```
ALTER TABLE fact100
    MODIFY COLUMN product_id int;
```

Import CSV files with tsload

Summary: Use the `tsload` command to import data from a CSV file.

Access: To add data into the system, you must have the required permissions. Initially, only the Admin user has that level of access.

Use ThoughtSpot Loader (`tsload`) to load data from a CSV text file into an existing table in ThoughtSpot. ThoughtSpot Loader (`tsload`) is a common way to import data. When using `tsload`, you can load larger datasets and make the loading process repeatable through scripting.

The `tsload` command accepts flags that enable you to specify column and row separators, date or timestamp formats, null value representations, and similar parameters. Many of these options have defaults that you can override.

Before importing data, you need to [build the schema \[See page 211\]](#).

After importing data, you can see it in the **Data** tab.



ThoughtSpot Loader

To use ThoughtSpot Loader, type the command `tsload` followed by the appropriate flags. You can see the list of the flags it accepts in the [ThoughtSpot Loader flag reference \[See page 0\]](#), or by issuing `tsload --help` on the command line.

The `tsload` feature supports both full and incremental data loads. For incremental loads, it performs an upsert (insert or update). If an incoming row has the same primary key as an existing row, it updates the existing row with new values.

You can integrate `tsload` into your ETL environment for more automated data loads. Most ETL tools enable you to write target data into files, and support scripted post-transformation actions that include loading data into ThoughtSpot.

Load data using `tsload`

To manually load data into ThoughtSpot, follow these steps.

You can also save the `tsload` commands to a script, for reuse.

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Change to the directory that contains the staged CSV files.

```
$ cd <directory>
```

3. Use the following syntax to invoke `tsload`, specifying the appropriate flags, and your data source file:

```
$ tsload --target_database=<my_database>
          --target_table=<my_table> --alsologtostderr
          --empty_target --source_file=<my_file.csv> --v 1
          --field_separator="separator_char"
```

This example imports the CSV file `ssbm_customer.csv` into the table CUSTOMER:

```
$ tsload --target_database=SAMPLE_DB
          --target_table=CUSTOMER --alsologtostderr
          --empty_target --source_file=ssbm_customer.csv
          --v 1 --field_separator "|"
```

- Once `tsload` starts processing the data, you can see messages that indicate the progress.

When the load is complete, you can see two summary messages: `source summary` and `load summary`.

```
Started processing data row
Source has 32 data rows, ignored row count 0
Waiting for rows to commit...(please wait)

Source summary
-----
Data source: ssbm_customer.csv
Source data format: CSV
Header row?: no
Tokenizer Options: escape_char: "" field
d_separator: "|" enclosing_char: "\\""
Date format: %Y%m%d
Date time format: %Y%m%d %H:%M:%S
Flexible mode?: no

Load summary
-----
Target table: CUSTOMER
Should empty target?: yes
Status: Successful
Rows total: 32
Rows successfully loaded: 30
Rows failed to load: 0
Rows duplicate/omitted: 2
% of Rows successfully loaded: 93.75 %
Load Rate (MB/s): 0.00 MB/s
Load Rate (Rows/s): 3.53 Rows/s
Start time (Wallclock): Tue Jan 29 09:09:07
End time (Wallclock): Tue Jan 29 09:09:08
Total load time = 1.13 seconds = 0.02 minutes = 0.00 hours
Data size = 50 bytes = 0.06 KB = 0.00 MB
```

- In the load summary, check the **Rows duplicate/omitted** number.

This indicates the number of rows (if any) that `tsload` did not load because they did not satisfy the table constraints. A common cause of this would be a duplicate primary key. If `tsload` omitted any rows, review your CSV file, make the required adjustments, and then load it again.

In the example above, `tsload` omitted two rows. This data needs to be cleaned and reloaded.

6. After correctly loading your file, repeat this process to load data from any additional CSV files.

Loading data from an AWS S3 bucket

If you have data in `.csv` format stored in an AWS bucket, you can load it directly to ThoughtSpot.

Assigning S3 read-only role to your EC2 instance

If your cluster is on version 5.3.1 or later, you can assign an S3 read-only role to your ThoughtSpot EC2 instance(s) so the instance(s) can access the S3 bucket from which you want to load the data. This eliminates the need to enter the AWS S3 credentials when loading your data. For details, see: [Using an IAM Role to Grant Permissions to Applications Running on Amazon EC2 Instances](#) (https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_use_switch-role-ec2.html) in Amazon's AWS documentation.

If you are using S3 for persistent storage, and assigned the `ec2rolewithfulls3access` IAM role to your instance, you do **not** need to complete this step in order to load data with S3.

If you already configured the IAM role, you do not need to supply the `--s3a_access_key`, `--s3a_secret_key` and `--s3a_region` parameters when loading data with `tsload`.

To load data from an AWS S3 bucket, do the following:

1. Log in to the Linux shell of your AWS instance using SSH.
2. Use the following syntax to invoke `tsload`, specifying the appropriate flags and your data source file. If you do not supply the `--s3a_bucket_name`, `tsload` prompts you to enter it.

```
$ tsload --source_file "/s3a/default/<my_file_in_aws>"  
    --target_database "<my_database_in_ThoughtSpot>"  
    --target_table "<my_table_in_the_database_in_ThoughtSpo  
t>" --s3a_bucket_name <bucket_name>
```

⚠ Important: `--source_file` should contain the file path inside the bucket and **must** be prefixed by `/s3a/default`. For example, to load from a file named “directory/file.csv”, the `--source_file` should be `"/s3a/default/directory/file.csv"`.

This example imports the CSV file `teams.csv` into the table `teams` in the database `temp`:

```
$ tsload --source_file "/aws/default/teams.csv"  
    --target_database "temp" --target_table "teams"
```

3. After running the `tsload` command, you are prompted to enter additional AWS S3 information:

- AWS S3a bucket name

If you have no IAM role configured, you must enter the following information as well:

- AWS S3a region
- AWS S3a access key
- AWS S3a secret key

Optionally, these four pieces of information can be inserted at the beginning of the command (in step 2), using the following flags:

- `--s3a_bucket_name "<Name of bucket that contains the source CSV
file>"`
- `--s3a_region "<Region where the bucket is located>"`
- `--s3a_access_key "<AWS S3 access key>"`
- `--s3a_secret_key "<AWS S3 secret key>"`

4. After the processing begins, progress messages appear, and then source and load summary messages after the load is complete.

Loading data from a GCP GCS bucket

If you have data in `.csv` format stored in a GCS bucket, you can load it directly to ThoughtSpot.

Assigning GCS read-only role to your GCP instance

You can assign a read-only role to your ThoughtSpot GCP instance(s) so the instance(s) can access the GCS bucket from which you want to load the data. This eliminates the need to enter the GCP GCS credentials when loading your data. For details, refer to the storage setting detailed in the GCP [Create an instance \[See page 0\]](#) section.

If you are using GCS for persistent storage, you already assigned the **Set access for each API** scope to your instance and specified **Full** storage access, so you do **not** need to complete this step in order to load data with GCS.

Create the database and table

1. Log in to the Linux shell of your GCP instance using SSH.
2. Invoke TQL:

```
$ tql
```

```
TQL>
```

3. Create the database:

```
TQL> CREATE DATABASE temp;
```

4. Connect to the database:

```
TQL> USE temp;
```

5. Create the table

```
TQL> create table teams (id int, name VARCHAR(255));  
TQL> exit;
```

Load data

Use the following syntax to invoke `tsload`, specifying the appropriate flags and your data source file:

```
$ tsload --source_file /gs/default/team.csv
      --target_database temp
      --target_table teams
      --bucket_name "my_gcs_bucket"
      --has_header_row 2>/dev/null

$ Header row read successfully
Source has 2 data rows, has header row, ignored row count 0
Waiting for rows to commit...(please wait)
Source summary
-----
Data source:                  /gs/default/team.csv
Source data format:          csv
Header row?:                 yes
Tokenizer Options:            escape_char: "" field_separator:
",," enclosing_char: "\" null_value: "(null)" trailing_field_se
parator: false
Date format:                  %Y%m%d
Date time format:            %Y%m%d %H:%M:%S
Flexible mode?:              no
Load summary
-----
Target table:                 teams
Should empty target?:        no
Status:                       Successful
Rows total:                   2
Rows successfully loaded:    2
Rows failed to load:         0
% of Rows successfully loaded: 100.00 %
Load Rate (MB/s):           0.00 MB/s
Load Rate (Rows/s):          1.13 Rows/s
Start time (Wallclock):       Wed Oct 30 23:30:11
End time (Wallclock):         Wed Oct 30 23:30:13
Total load time = 1.78 seconds = 0.03 minutes = 0.00 hours
Data size = 19 bytes = 0.02 KB = 0.00 MB
```

Verify the table contents

1. Invoke TQL:

```
$ tql
```

```
TQL>
```

2. Connect to the database:

```
TQL> USE temp;
```

3. Show the table data:

```
TQL> select * from teams;
```

Table contents are displayed:

id	name
1	sameer
2	sandeep

Use a script to load data

Summary: Learn how to load files in bulk with a script.

If you need to load data from multiple CSV files, create a script to automate the process. You can also use a similar script to automate recurring data feeds.

Understand how to create a script

The data loading script is a text file that contains all the calls to `tsload` for loading the data from your CSV files.

The example script shown here uses the `cat` command to read the data file, and pipes it to `tsload`. When creating and testing your script, you may wish to replace each `cat` with `cat -10`, to load only the first ten lines of each file. This allows you to quickly run a test of your script. When the test succeeds for all the data files, you can then remove each `-10`, so the complete files will load when you run the script again.

1. Log in to the Linux shell using SSH.
2. Navigate to the directory that contains your CSV files and open a new file in a text editor.
3. Type in the commands to load the data.

This example shows commands to load three files:

```
cat Players.csv | tsload
  --target_database baseball --target_table "players"
  --empty_target --field_separator ","
  --max_ignored_rows 10 --bad_records_file bad_record
s.txt
  --has_header_row --alsologtostderr --null_value ""

cat AllstarFull.csv | tsload
  --target_database baseball --target_table "allstarfu
ll"
  --empty_target --field_separator ","
  --max_ignored_rows 10 --bad_records_file bad_record
s.txt
  --has_header_row --alsologtostderr --null_value ""

cat Appearances.csv | tsload
  --target_database baseball --target_table "appearanc
es"
  --empty_target --field_separator ","
  --max_ignored_rows 10 --bad_records_file bad_record
s.txt
  --has_header_row --alsologtostderr --null_value ""
```

4. Save the file.

5. Run the script:

```
$ ./load_baseball_data.sh
```

Loading data efficiently

If you have a very large data file that takes a long time to load, you can reduce the load time by splitting it up into multiple files and loading them in parallel using multiple invocations of `tsload`. If the size of any of your data files is greater than 50 million rows, running `tsload` in parallel can reduce the load time significantly.

- Split up your large data file into multiple smaller files.
- Stage the data files in a location accessible to the node on which you run the script. Usually, you can use an [NAS mounted file system](#) [See page 132].

- Create a script to load the files in the following example.
- Run the script to load the files. You will make your script multi-threaded by invoking multiple loader threads (between 1 and 5 are recommended).

To optimize the load time even further, determine what the bottleneck is and adjust your process accordingly. If the disk I/O for reading the data files is the bottleneck, you can stage the data files on separate NAS mounted file systems and reference them accordingly in your script. If the CPU on the machine you're using to run the load script is the bottleneck, you can split the load script into the same number of parts as you have nodes in your ThoughtSpot instance, place one script on each node, and run them in parallel. Make sure the other nodes are able to access the data files where they are staged. Running the load script on separate nodes will put the data on all the nodes, just as when you run the script on a single node. Running the script on all the nodes at the same time just lets you take advantage of CPU power of each node for hashing data files.

For example, suppose you have 30 days of data in 30 files, one for the data collected on each day. Each day's data file contains 10 million rows, for a total of 300 million rows of data. You want to load the whole month of data. For this example we'll have 5 loader processes, each one handling 6 days of data. Here is a sample script you could use to load the data files in parallel:

```
/* Script load_script.sh, loads 30 days of data in parallel */

#!/bin/bash

pidlist=""

cat day1.csv day2.csv day3.csv day4.csv day5.csv day6.csv | tsload
--target_database sales --target_table SALES_FACT --max_ignore
d_rows 10
--bad_records_file ./SALES_FACT.bad --date_format %Y-%m-%d
--date_time_format "%Y-%m-%d %H:%M:%S" --source_data_format del
imited
--field_separator "|" --null_value "" --enclosing_character
"\\""
--boolean_representation 1_0 &

pidlist="$pidlist $" &

cat day7.csv day8.csv day9.csv day10.csv day11.csv day12.csv | tsload
--target_database sales --target_table SALES_FACT --max_ignore
d_rows 10
--bad_records_file ./SALES_FACT.bad --date_format %Y-%m-%d
--date_time_format "%Y-%m-%d %H:%M:%S" --source_data_format del
imited
--field_separator "|" --null_value "" --enclosing_character
"\\""
--boolean_representation 1_0 &

pidlist="$pidlist $" &

cat day13.csv day14.csv day15.csv day16.csv day17.csv day18.cs
v | tsload
--target_database sales --target_table SALES_FACT --max_ignore
d_rows 10
--bad_records_file ./SALES_FACT.bad --date_format %Y-%m-%d
--date_time_format "%Y-%m-%d %H:%M:%S" --source_data_format del
imited
--field_separator "|" --null_value "" --enclosing_character
"\\""
--boolean_representation 1_0 &

pidlist="$pidlist $" &
```

```
cat day19.csv day20.csv day21.csv day22.csv day23.csv day24.cs
v | tsload
--target_database sales --target_table SALES_FACT --max_ignore
d_rows 10
--bad_records_file ./SALES_FACT.bad --date_format %Y-%m-%d
--date_time_format "%Y-%m-%d %H:%M:%S" --source_data_format del
imited
--field_separator "|" --null_value "" --enclosing_character
"\\""
--boolean_representation 1_0 &

pidlist="$pidlist $" &

cat day25.csv day26.csv day27.csv day28.csv day29.csv day30.cs
v | tsload
--target_database sales --target_table SALES_FACT --max_ignore
d_rows 10
--bad_records_file ./SALES_FACT.bad --date_format %Y-%m-%d
--date_time_format "%Y-%m-%d %H:%M:%S" --source_data_format del
imited
--field_separator "|" --null_value "" --enclosing_character
"\\""
--boolean_representation 1_0 &

pidlist="$pidlist $" &

wait $pidlist
```

Call your script using a command like:

```
nohup bash ./load_script.sh > master_log.txt &
tail -f master_log.txt
```

Constructing your script in this way will execute all the commands in the background, and output to the file `master_log.txt`. You can see a running status as the commands in the script execute. After the script completes, you can check the log file for detailed information, such as the number of rows that loaded successfully.

Use the tsload connector to load data

Summary: Learn how to load files using the tsload connector.

Another option for loading data in bulk, is to use tsload connector. tsload connector is a collection of APIs.

This page highlights the following:

- Setting up your cluster to use tsload connector
- Using the reference client for writing automated ETL jobs
- Server and authentication details
- API workflow inside the client

Setting up your cluster

1. SSH as admin into your ThoughtSpot cluster: `ssh admin@<cluster-ip-address or hostname>`.
2. Open the config file at the following location: `/usr/local/scaligent/release/production/orion/etl_http_server/prod.config`
3. If your cluster is behind a load balancer, you must disable the internal etl server's load balancer. Contact ThoughtSpot support for assistance with this step.
4. By default, bad-records are saved in one of the mounted drives. If that is not possible, they are saved to `/tmp`. To modify this location, contact ThoughtSpot support.
5. If your cluster has been upgraded from an earlier version, validate that your SSL certificates are bound to tsload connector. Contact ThoughtSpot support for assistance with this step.

Note: For clusters created in version 6.2 or later, certificates are bound to tsload connector by default.

6. Check if the **etl_http_server**, responsible for tsload connector, is accessible by pinging it:

```
curl -i https://localhost:8442/ts_dataservice/v1/public/  
ping  
HTTP/1.1 200 OK
```

Copy ?

Ping Received.

Reference client

The included Python3 client is available for you to use it as a starting point for writing automated ETL jobs in Python. Contact ThoughtSpot for more information.

This requires python3 and the details of the methods are documented within the above client.

The client includes the following methods:

- **login**: Requires ThoughtSpot username and password
- **startLoad**: Prepares the load with parameters that include table, schema, and db. It also includes tokenizing parameters, like : field separator etc.
 - This returns a new IP address (when the internal load balancer is enabled)
 - This also returns a `cycle_id`. This `cycle_id` determines the load-session corresponding to the given load parameters. All the successive calls will need to use this `cycle_id` as a parameter, including the `getStatus`.
 - LoadDataParam JSON

```
{  
  "target": {  
    "database": "mytest",  
    "schema": "falcon_default_schema",  
    "table": "one"  
  },  
  "format": {  
    "date_time": {  
      "date_format": "%Y%m%d",  
      "time_format": "%H-%M-%S",  
      "date_time_format": "%Y%m%d %H:%M",  
      "converted_to_epoch": false,  
      "second_fraction_start": ".",  
      "skip_second_fraction": false  
    },  
    "boolean": {  
      "true_format": "T",  
      "false_format": "F",  
      "use_bit_values": false  
    },  
    "type": "CSV",  
    "field_separator": ",",  
    "trailing_field_separator": true,  
    "enclosing_character": "^",  
    "escape_character": "^",  
    "has_header_row": true,  
    "null_value": "(null)",  
    "flexible": false  
  },  
  "load_options": {  
    "empty_target": true,  
    "max_ignored_rows": 10  
  }  
}
```

- **load:** in this example, a file is being uploaded in a single call. In reality, this could be a post-call with data directly instead of a file.
 - This can also be broken into multiple load calls for load data incrementally.
 - This simply uploads and starts processing the file, but the load is not complete just by calling this method.
 - This method returns immediately, and the actual parsing, etc is done asynchronously.
 - To get the status at any point, the **getStatus** method can be used.

- **commitLoad:** This method commits the current ingested data into the Falcon database.
 - This method returns immediately, and the commit is done asynchronously.
 - Again calling the **getStatus** method can provide the actual status.
- **getStatus:** Returns the status of the load at that time
 - **getStatus JSON**

```
{  
    "buffered_data": "0 Bytes",  
    "cycle_id": "78aecb14-34c5-4da8-b08f-517d  
e22d9701",  
    "end_time": "Mon, 08 Jun 2020 20:46:21 IS  
T",  
    "ignored_row_count": "0",  
    "ingested_network_bw": "0 Bytes",  
    "internal_stage": "DONE",  
    "rows_written": "2",  
    "size_written": "34 Bytes",  
    "start_time": "Mon, 08 Jun 2020 20:46:19  
IST",  
    "status": {  
        "code": "OK"  
    }  
}
```

Server and authentication details

Ports and Server

Port number: 8442, HTTPS REST endpoints

Note: Port 8442 is open by default in ThoughtSpot release 6.1 or later.

The load server resides on a different port compared to standard ThoughtSpot services. This is because the service tends to carry heavy file-load operations, and having a separate web server creates the needed isolation between standard ThoughtSpot services and tsload operations.

By default, this service runs on all nodes of a ThoughtSpot cluster. This provides load distribution to address possible simultaneous loads. The tsload server uses its own load balancer. If an external load balancer is used, the tsload requests must be sticky, and the tsload load balancer should be disabled.

Authorization and Authentication

This uses the existing ThoughtSpot authentication mechanism to authenticate the user, using the **Login** API. Each upload session must be authenticated using this API.

tsload is available only to users who have the “Administrator” or “Manage Data” privilege in the ThoughtSpot environment.

API workflow

The typical workflow of the API inside the client is the following:

1. `<standard-thoughtspot-cluster-url> Login .`
2. `<standard-thoughtspot-cluster-url> StartLoad .` If the tsload-LoadBalancer is turned on, this returns the new IP address (for one of the nodes in the cluster).
3. `<thoughtspot-node-ipReturnedFrom2> Load .`
 - a. Repeat this step until all the rows are sent.
 - b. In the case of a file, you can call this in one operation. In the case of a stream, you call this multiple times, thus avoiding buffering large data on the client side.
4. `<thoughtspot-node-ipReturnedFrom2> EndLoad .`
 - a. This will start the commit process.
 - b. It'll take some time for the data to be committed to Falcon Database.
5. `<thoughtspot-node-ipReturnedFrom2> GetStatus .`
 - a. To monitor the state of the commit.

- b. Wait until it returns “DONE”.

Related information

[tsload connector API reference \[See page 0\]](#)

Delete a data source

Summary: How to prepare for and delete a data source using the ThoughtSpot application.

There are two separate ways to delete a data source, through the browser or [through TQL](#) [See page 262] describes the dependency checking that occurs when deleting or changing a table using TQL.

When you want to delete a data source, you first need to handle any dependent objects that have been built on top of it. You can easily see these dependencies, and choose how to handle them before deleting the data source.

Check data source dependencies

You can see all of the dependencies for any data source (worksheet or table) on the **Data** page.

To view dependent objects for a data source:

1. Click **Data** on the top navigation bar.
2. Click the name of the data source whose dependencies you want to view.
3. Click **Dependents**.

You will see a list of the names of the dependent objects (worksheets and pinboards), and the columns they use from that data source. You can use this information to determine the impact of changing the structure of the data source or to see how widely it is used.

The screenshot shows a 'WORKSHEET' titled 'Sporting Goods Retail Worksheet' under 'ThoughtSPORT worksheet'. The 'Dependents' tab is selected. A table displays the following data:

COLUMN NAME	DEPENDENT NAME	TYPE
Age Group	Total Sales by Depar..	View
Date	Top 100 Products M..	View
Product Name	Top 100 Products M..	View
Department	Total Sales by Depar..	View
Sales	Total Sales by Depar..	View
Customer City	Customer Location	Answer
Customer Name	Customer Location	Answer

(showing rows 1-0 of 27)

4. Click a dependent object to modify or delete it.

If you want to remove the dependency by modifying the dependent object, you must remove all search terms or columns that refer back to the data source you are trying to delete.

5. When all dependencies have been removed, you will be able to go back and delete the data source.

Delete a data source

You can delete data sources from the browser, as long as they were not created by an administrator through `tsload`.

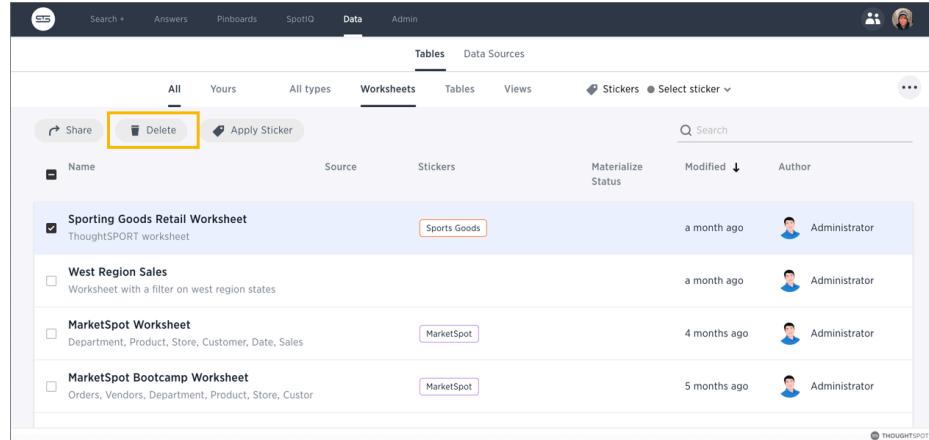
You can delete data sources from the browser if they were created in the browser. These types of data sources include:

- Data imported from the browser.
- Worksheets.

ThoughtSpot checks for dependencies whenever you try to delete a table or worksheet.

1. Click **Data** on the top navigation bar.
2. Check the box next to the name of the data source you want to delete.
3. Click the delete icon.

Delete a data source



The screenshot shows the ThoughtSpot interface with the 'Data' tab selected. Under the 'Worksheets' tab, there is a list of four worksheets. The first worksheet, 'Sporting Goods Retail Worksheet', has a checkbox next to its name. A yellow box highlights the 'Delete' button, which is located in the toolbar above the list. The other three worksheets listed are 'West Region Sales', 'MarketSpot Worksheet', and 'MarketSpot Bootcamp Worksheet'. Each worksheet entry includes a checkbox, a 'Source' column, a 'Stickers' column, a 'Materialize Status' column, a 'Modified' column, and an 'Author' column.

Name	Source	Stickers	Materialize Status	Modified	Author
Sporting Goods Retail Worksheet ThoughtSPORT worksheet	Sports Goods			a month ago	Administrator
West Region Sales Worksheet with a filter on west region states				a month ago	Administrator
MarketSpot Worksheet Department, Product, Store, Customer, Date, Sales	MarketSpot			4 months ago	Administrator
MarketSpot Bootcamp Worksheet Orders, Vendors, Department, Product, Store, Custor	MarketSpot			5 months ago	Administrator

4. If you attempt to delete a data source with dependent objects, the operation will be blocked.

You will see a list of dependent objects with links.

Cannot delete

The following object(s) depend on "Sporting Goods Retail Worksheet". You must delete them to delete "Sporting Goods Retail Worksheet".

- [Sales by Store - Last 30 Days](#) (Answer)
- [Shopping Trend of Members vs Non-Members](#) (Answer)
- [Sales Breakdown by Mode of Payment](#) (Answer)
- [Moving Sum of All Sales - Last 30 days](#) (Answer)
- [Pivot Example](#) (Answer)
- [Sales Trend by Day of Week](#) (Answer)
- [Customer Location](#) (Answer)
- [Customer Footprint and Sales by Region](#) (Answer)
- [Sales by Region, State and Year](#) (Answer)
- [Low Inventory](#) (Answer)
- [Monthly Department Sales Analysis](#) (Answer)
- [Sales for Last Month](#) (Answer)
- [Product & Department Sales Group Sum](#) (Answer)
- [Margin vs Sales Analysis](#) (Answer)
- [Sales by Age Group, Gender and Product Category](#) (Answer)
- [Racquet Sales](#) (Answer)
- [Sales by Quarter](#) (Answer)
- [Average Sales - Weekday vs Weekend](#) (Answer)
- [Vicky's Sales Data](#) (Answer)
- [Sales Per Customer for Outerwear by State](#) (Answer)

OK

5. Click a dependent object to modify or delete it.

If you want to remove the dependency by modifying the dependent object, you must remove all search terms or columns that refer back to the data source you are trying to delete.

6. When all dependencies have been removed, you will be able to go back and delete the data source.

Delete or change a table in TQL

Summary: You can delete a data source in the web browser or using ThoughtSpot SQL Command Line (TQL).

When you enter a TQL statement, the system warns you of possible dependency consequences with a prompt asking if you'd like to proceed. This should make you feel safe issuing TQL commands, even commands like dropping a table.

TQL actions with possible dependency consequences include:

- Change, add, or remove a primary key.
 - When changing or adding a primary key, if the key in question is not unique in the data it may cause deletion of rows, because of upserts occurring when duplicate primary keys are found.
 - When changing or removing a primary key, incoming foreign key relationships will be broken.
- Change a column datatype.
- Add a relationship or foreign key.
- Drop a relationship or foreign key constraint.
- Change or add a sharding key.
- Drop a table, schema, or database.

When issuing one of the previous commands, you will see a warning message similar to this:

```
TQL> ALTER TABLE table1
      DROP CONSTRAINT PRIMARY KEY;

WARNING: This operation will break the Foreign Key relationship "products"
with table "sales", which will break 34 user-visible visualizations and
2 Worksheets. We recommend taking a snapshot before performing
this operation.
Do you wish to proceed? (yes/no).
```

Onboarding Users

Summary: Guided onboarding simplifies the initial engagement that new users have with ThoughtSpot, and encourages adoption throughout your organization.

Onboarding enables users to master the key workflows of ThoughtSpot, and makes them productive and capable users in a very short time.

When you create a new user, we recommend that you add them to a user group immediately. Configure that user group to use a specific data source, choose up to three initial pinboards, and specify the text of the welcome email.

Prerequisites for onboarding

- **Valid emails** All users must have valid emails.

To include users in the onboarding process, each user profile must include a valid email address; see [Create a user \[See page 287\]](#). Contact ThoughtSpot Support [\[See page 0\]](#) to add new email domains to your list of allowed domains.

To load and validate user information (including email addresses) in bulk, we recommend using Active Directory configuration and sync. See [Configure LDAP for Active Directory \[See page 118\]](#).

- **Onboarding configuration** Before starting the onboarding process, an administrator must specify the configuration through the `tscli onboarding` command [\[See page 0\]](#).

Onboarding Process

The key workflows that enable successful onboarding include the following:

1. Configure and save a default Pinboard (or Pinboards) for new users. Alternatively, determine what default Pinboards you plan to use for each user group.

See steps for [creating a pinboard \[See page 0\]](#).

2. Create a new user group, or edit an existing user group to which new users belong:
 - specify the group name and its display name
 - set the sharing visibility to visible
 - specify the Privileges
 - select up to 3 default pinboards
 - test and customize the test welcome email. See steps for [creating a user group \[See page 271\]](#) or [editing a user group \[See page 275\]](#).
3. Create a new user and assign them to the group you are using for onboarding. See steps for [creating a new user \[See page 287\]](#). You can also add existing users to the group; see [In addition to creating a new user directly in ThoughtSpot, you can add users in bulk through Active Directory \[See page 115\]](#).
4. When the new user signs in for the first time, the ThoughtSpot onboarding process starts automatically and guides them through a few basic scenarios of using the application. See [User onboarding experience \[See page 0\]](#).
5. **[Optional]** Any user can repeat their onboarding experience at any time. Simply select **Profile** from user icon on the top right corner of the page. Under **Preferences > New user onboarding**, click **Revisit**. See [Revisit onboarding \[See page 0\]](#).

Notes on Data Source recommendation

User groups in ThoughtSpot can be hierarchical, and each user can belong to multiple groups. Because of this, the choice of the Recommended Data Source may be surprising, both to users and admins.

We determine the default Data Source for **each user** based on these criteria:

- Consider all default Pinboards, across all Groups
- Identify Data Sources associated with these Pinboards
- Rank the Data Sources on frequency of use
- Select the highest ranked Data Source

Admins can preview onboarding flow for any given user, and adjust the selection of default Pinboards.

Understand groups and privileges

Summary: Creating groups and assigning users to them makes privilege management easier.

Before people can log in and use ThoughtSpot, you need to create a username, a password, and a membership in one or more groups for them.

This page describes manual creation of users, groups, and privileges, but you can also manage users through [LDAP \[See page 117\]](#) or SAML. For information on configuring SAML authentication, see [Configure SAML \[See page 109\]](#).

Privileges and groups

Privileges determine what kinds of actions users are allowed to do. You assign privileges to groups. Then, you create users and assign them to groups. This is how you grant users access to different capabilities in ThoughtSpot.

Each group includes a set of privileges for its users. The privileges a group has determine the actions that its members are allowed to do. If a user belongs to more than one group, they will have the highest level of the privileges from all the groups they belong to. Plan your groups so that you can use them to assign a common set of privileges to multiple users. Good planning will pay off in ease of administration and a better search experience.

There is a default group called **All**, which includes every user in ThoughtSpot. When you create a new user, they will be added to the **All** group automatically. You cannot delete the **All** group or remove members from it.

You can also have a hierarchy of groups. That is, groups can belong to (that is, be children of) other groups. When using group hierarchies, permissions are inherited from the parent group. So if you're a member of a sub-group, you would automatically have the privileges of the parent group.

List of privileges

Here are the different privileges, and the capabilities they enable:

Can administer ThoughtSpot

Can manage Users and Groups and has view and edit access to all data. Users with this privilege can also download a saved Answer.

Can upload user data

Can upload their own data from the application's **Data** page using **Actions > Upload data**.

Can download data

Can download data from search results and pinboards.

Can share with all users

Can see the names of and share with users outside of the groups the user belongs to. Members of groups with this privilege can also share with groups marked as **NOT SHARABLE**.

Can manage data

Can create worksheets and views. Can create Embrace connections.

Note that to edit a worksheet or a view created by another user, you must have the **Edit** permission on that object, and it must be shared with you.

Can use experimental features

Can access trial and experimental features that ThoughtSpot makes available to early adopters.

Can invoke Custom R Analysis

Can access R scripts to further explore search answers. Includes options to invoke R scripts on visualizations, create and share custom scripts, and share the results of R analysis as answers and pinboards.

Can schedule pinboards

Can create pinboard schedules and edit their own scheduled jobs.

Has SpotIQ privilege

Can use the SpotIQ feature.

If this privilege is not enabled for the user, but Insights are enabled on the cluster (this is off by default), they can still see "Did you know" SpotIQ insights on the ThoughtSpot home page.

Can administer and bypass RLS

Users in groups with this privilege (directly or through group inheritance):

- Are exempt from row-level security (RLS) rules.
- Can add/edit/delete existing RLS rules.
- Can check or uncheck Bypass RLS on a worksheet.

Your installation configuration may enable or disable the availability of this privilege. By default, it is enabled. Administrators or groups with the privilege **Can administer ThoughtSpot** can grant this privilege.

Cannot create or update Pinboards

Users are limited to viewing and exploring curated Pinboards (and Answers). They cannot copy, edit, download, or share them.

This privilege is designed to support embedded implementations, and it is behind a flag.

See [Granular access to Pinboards \[See page 0\]](#) for a deeper discussion of this privilege, and an implementation example.

Typically, the **ALL** group has a common set of privileges applies such as the **Can upload user data** and/ or **Can download data** privileges.

Privileges are additive, meaning that if a user belongs to more than one group, they will have the highest level of privileges from among the groups they are a member of. They are also inherited from the parent, so that a sub-group gets all the same privileges of its parent, all the way up the group hierarchy.

If you add the privilege **Has administration privileges** to a group, note that all users in that group will be able to see all the data in ThoughtSpot. Administrators can see all data sources, and [Row level security \[See page 349\]](#) does not apply to them.

Permissions to see and edit tables, worksheets, and pinboards are set when you share them with users and groups, as described in the topic [Data security \[See page 308\]](#).

The following table shows the intersection of user privilege and ability:

	Create/Edit WS	Create View	Create Embrace Connection	Modify Col. Props. ¹	Upload Data	Download Data	Share within Group	Share with all users	Manage RLS rules	CrUD Relationships	Read Relationships	See Hidden Cols	Join with Upload Data	Schema Viewer	Use Scheduler	Use Auto-Analyze
Can administer ThoughtSpot	Y	Y	Y	Y	Y	Y	Y	Y	Y ²	Y	Y	Y	Y	Y	Y	Y
Can upload user data	N	N	N	N	Y	Y	Y	Y	Y ³	Y ⁴	N	N	N	N	N	N
Can download data	N	N	N	N	N	Y	Y	N	N	Y ⁴	N	N	N	N	N	N
Can manage data	Y	Y	Y	Y	Y	Y	N	Y	Y ⁴	Y ⁴	Y ⁵	Y	N	N	N	N
Can share with all users	N	N	N	N	N	Y	Y	N	N	Y ⁴	N	N	N	N	N	N
Has SpotIQ privilege	N	N	N	N	N	N	N	N	N	Y ⁴	N	N	N	N	N	N
Can Administer and Bypass RLS	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
None	N	N	N	N	Y	N	N	Y ⁴	N	N	N	N	N	N	N	N

Table notes:

1. Applies to non-owners only.
2. Any tables.
3. Author of at least one table in relationship.
4. Only when read permission for columns used in the relationship.
5. With edit permission.



Related information

- [Add a group and set security privileges \[See page 271\]](#)
- [Add a user \[See page 287\]](#)

Create, edit, or delete a group

Summary: ThoughtSpot has intuitive and powerful user group management for assigning privileges, user selection, multi-tier subgroups, default Pinboard assignment, and emailing.

Before adding users, create the groups to which they belong. Each group includes a set of privileges for its users.

Create a group

To create a group and add privileges for the group, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Groups** from the side navigation bar that appears.

A screenshot of the ThoughtSpot Admin Console Groups page. On the left, there is a sidebar with various administrative sections like User Management, Authentication, Settings, and System. The "Groups" section is currently selected and highlighted with a blue box. The main content area shows a table of existing groups. The columns are labeled "Display Name", "Name ↑", "Created", and "Default pinboards". There are two entries: "Administration Group" (display name A, created 8 years ago, 0 pinboards) and "Group" (display name G, created 21 hours ago, 0 pinboards). At the top left of the main content area, there is a blue button labeled "+ Add Group". The top navigation bar is visible at the very top of the page.

4. Click the **+ Add Group** button on the upper left-hand side of the list of groups.

5. In the **Add a new group** modal, enter the details for the new group:

Field	Description
Group name	Enter a unique name for the group.
Display name	Name of the group as it appears in ThoughtSpot.

Sharing visibility	Indicate whether objects can be shared with this group. When set to SHARABLE , this group is an option in the Share dialog.
Description	Optionally, enter a description.
Privileges	Check the privileges [See page 265] you want to grant to the group. If you add the privilege Has administration privileges to a group, all users in that group can see all the data in ThoughtSpot. Administrators can always see all data sources, and Row level security [See page 349] does not apply to them.

6. You can also add [Groups \[See page 274\]](#) (these would be the subgroups of the group you are editing), [Users \[See page 274\]](#). Also, see [Default Pinboards \[See page 0\]](#).
7. Click **Add** to create the group.

Default Pinboards

You cannot add default Pinboards to a new user group. You must create it first, and then edit it to add default Pinboards. See [Edit a group > Default Pinboards \[See page 277\]](#).

Add a new group

Group name *	Sales EMEA
Display name *	Sales EMEA
Sharing visibility *	SHARABLE
Description	
Privileges	<input type="checkbox"/> Can administer ThoughtSpot <input checked="" type="checkbox"/> Can upload user data <input checked="" type="checkbox"/> Can download data <input type="checkbox"/> Can share with all users <input type="checkbox"/> Can manage data <input type="checkbox"/> Can use experimental features <input type="checkbox"/> Can invoke Custom R Analysis <input type="checkbox"/> Has SpotIQ privilege <input type="checkbox"/> Can administer and bypass RLS

Select default pinboards [i](#)

Steps to setup default pinboards for this group:

1. Create this group
2. Share existing or new Pinboards with this group
3. Edit this group and assign default Pinboards

Cancel ADD

* Required field

Groups

Follow these steps to assign subgroups to the group:

1. Click the **Groups** tab.
2. Select the groups you want to add in the list by clicking the box next to the group name.
3. You can also use **Search** to find groups by name.

Add a new group

Group name *	Sales EMEA
Display name *	Sales EMEA
Sharing visibility *	SHARABLE
Description	
Privileges	<input type="checkbox"/> Can administer ThoughtSpot <input checked="" type="checkbox"/> Can upload user data <input checked="" type="checkbox"/> Can download data <input type="checkbox"/> Can share with all users <input type="checkbox"/> Can manage data <input type="checkbox"/> Can use experimental features <input type="checkbox"/> Can invoke Custom R Analysis <input type="checkbox"/> Has SpotIQ privilege <input type="checkbox"/> Can administer and bypass RLS

No Groups in Group

Q Search by name

Sales Engineering
 Sales Executives
 Sales Reps

Cancel **ADD** 

Users

Follow these steps to assign users to the group:

1. Click the **Users** tab.
2. Select the users you want to add in the list by clicking the box next to the user name.
3. You can also use **Search** to find users by name.

Add a new group

Group name *

Display name *

Sharing visibility * SHARABLE

Description

Privileges Can administer ThoughtSpot
 Can upload user data
 Can download data
 Can share with all users
 Can manage data
 Can use experimental features
 Can invoke Custom R Analysis
 Can schedule pinboards
 Has SpotIQ privilege
 Can administer and bypass RLS

No User in Group

Q Search by name

Clear all | Select all

john@thoughtspot.com
 paul@thoughtspot.com
 tsadmin

* Required field

Edit a group

After adding a group, you can edit its settings to add or revoke privileges. The new settings apply to all group members.

To edit an existing group, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Groups** from the side navigation bar that appears.

Display Name	Name ↑	Created	Default pinboards
Administration Group	Administrator	8 years ago	0
Group	Group	21 hours ago	0

- Find the group you want to edit in the list, and click its name.

If you don't immediately see the name of the group, try searching for it.

- You can change the [Group name](#) [See page 272], [Display name](#) [See page 272], [Sharing visibility](#) [See page 0], [Description](#) [See page 273], and the selected [Privileges](#) [See page 273].

Here, we added the *Can manage data* privilege.

Edit group

Group name *	Sales EMEA
Display name *	Sales EMEA
Sharing visibility *	SHARABLE
Description	
Privileges	<input type="checkbox"/> Can administer ThoughtSpot <input checked="" type="checkbox"/> Can upload user data <input checked="" type="checkbox"/> Can download data <input type="checkbox"/> Can share with all users <input checked="" type="checkbox"/> Can manage data <input type="checkbox"/> Can use experimental features <input type="checkbox"/> Can invoke Custom R Analysis <input type="checkbox"/> Has SpotIQ privilege <input type="checkbox"/> Can administer and bypass RLS

* Required field

Default Pinboards Groups Users Email

Select default pinboards *i*

Search by name

- AE Pinboard
- Activity Dashboard
- Campaign Dashboard
- Deal Flow Analysis
- Demo - Customer Sales Metrics
- Free Trial Usage Analysis
- Marketing Attribution
- Marketing Demand Generation
- Marketing Funnel

Cancel **Update** 

You can also make changes to the [Default Pinboards \[See page 277\]](#), [Groups \[See page 278\]](#) (these would be the subgroups of the group you are editing), [Users \[See page 279\]](#), or [Email \[See page 280\]](#).

6. Make your changes, and click **Update**.

Default Pinboards

To assign default pinboards to groups, follow these steps:

1. Create a group, or choose an existing group. Note its name.
2. In the **Pinboards** interface, find the correct Pinboards, and share them with this group. See [Share a Pinboard \[See page 326\]](#).
3. Open the Group for editing. See [Edit a group \[See page 275\]](#).
4. Click the **Default Pinboards** tab.
5. From the list of shared Pinboards, select 1-3 default Pinboards in the list by clicking the box next to the Pinboard name.

6. You can also use **Search** to find Pinboards by name.
7. Click **Update** to save changes.

Groups

When editing a group, keep in mind that only subgroups or possible subgroups appear in the list of groups. The **No Groups in Group** only indicates there are no children in this group's hierarchy. Do not underestimate the importance of the parent(s) of the group, because each group inherits the privileges of each of its parent groups.

Follow these steps to change subgroups of the group:

1. Click the **Groups** tab.
2. Select the groups you want to add in the list by clicking the box next to the group name.
3. You can also use **Search** to find groups by name.
4. Deselect the groups you want to remove from the list by clearing the box next to the group name.
5. Click **Update** to save changes.

Edit group

Group name * Sales EMEA

Display name * Sales EMEA

Sharing visibility * SHARABLE

Description

Privileges

- Can administer ThoughtSpot
- Can upload user data
- Can download data
- Can share with all users
- Can manage data
- Can use experimental features
- Can invoke Custom R Analysis
- Has SpotIQ privilege
- Can administer and bypass RLS

* Required field

Cancel Update

Default Pinboards Groups Users Email

No Groups in Group

Search by name

Clear all Select all

- Sales Engineering
- Sales Executives
- Sales Reps

Users

Follow these steps to change the users of the group:

1. Click the **Users** tab.
2. Select the users you want to add in the list by clicking the box next to the user name.
3. You can also use **Search** to find users by name.
4. Deselect the users you want to remove from the list by clearing the box next to the user name.
5. Click **Update** to save changes.

Edit group

Group name * Sales EMEA

Display name * Sales EMEA

Sharing visibility * SHARABLE

Description

Privileges

- Can administer ThoughtSpot
- Can upload user data
- Can download data
- Can share with all users
- Can manage data
- Can use experimental features
- Can invoke Custom R Analysis
- Has SpotIQ privilege
- Can administer and bypass RLS

* Required field

Default Pinboards Groups **Users** Email

11 Users in Group

Search by name

Clear all | Select all

- user_sharing_82
- user_sharing_83
- user_sharing_84
- user_sharing_85
- user_sharing_86
- user_sharing_87
- user_sharing_88
- user_sharing_89
- user_sharing_9
- user_sharing_90
- user_sharing_91

Cancel Update

Email

You can configure groups so that users receive a *welcome email* that introduces them to ThoughtSpot, and initiates the onboarding process.

Follow these steps to configure group-wide emails:

1. Click the **Email** tab.
2. Under **Resend welcome email**, select either either *All users* or *New users*.
3. Enter optional text for the email. Here, we added “Welcome!”
4. To send the email immediately, click **Send**.
5. To test the email, click “Test welcome email”
6. Click **Update** to save changes.

Edit group

Group name * Sales EMEA

Display name * Sales EMEA

Sharing visibility * SHARABLE

Description

Privileges

- Can administer ThoughtSpot
- Can upload user data
- Can download data
- Can share with all users
- Can manage data
- Can use experimental features
- Can invoke Custom R Analysis
- Has SpotIQ privilege
- Can administer and bypass RLS

* Required field

Cancel Update

Default Pinboards Groups Users Email

Resend welcome email

All users New users

Welcome!

Send

Test welcome email

Deleting groups

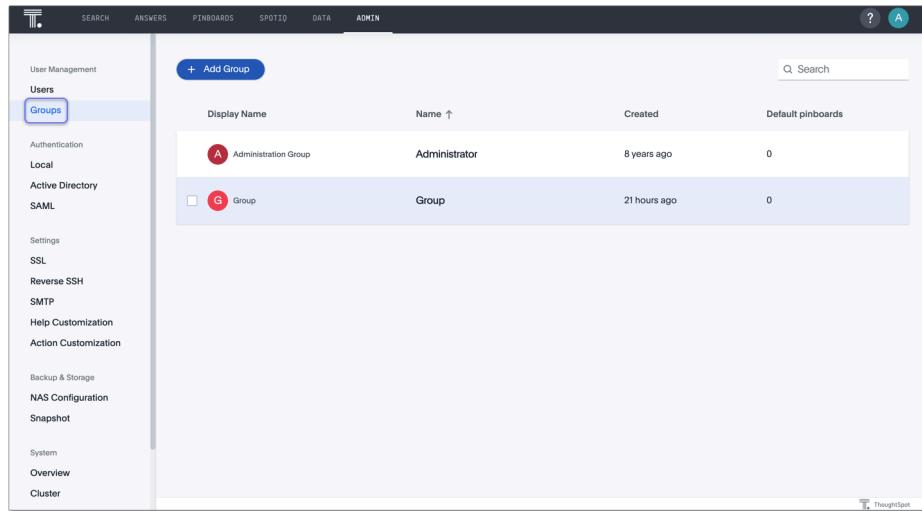
To delete existing groups, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Groups** from the side navigation bar that appears.

Create, edit, or delete a group

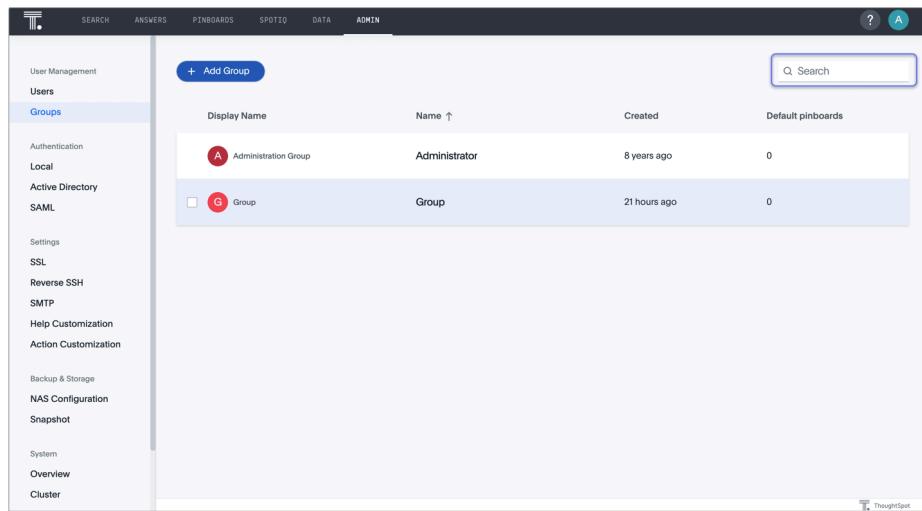


The screenshot shows the ThoughtSpot Admin interface with the 'Groups' tab selected in the sidebar. The main area displays a table of groups. The columns are 'Display Name', 'Name ↑', 'Created', and 'Default pinboards'. There are two entries: 'Administration Group' (created 8 years ago) and 'Group' (created 21 hours ago). A search bar at the top right is empty.

Display Name	Name ↑	Created	Default pinboards
A Administration Group	Administrator	8 years ago	0
G Group	Group	21 hours ago	0

4. Select the groups you plan to delete by clicking the box next to the group name.

If you don't immediately see the name of the group, try searching for it.



The screenshot shows the ThoughtSpot Admin interface with the 'Groups' tab selected in the sidebar. The main area displays a table of groups. A search bar at the top right is now highlighted with a blue border. The table has the same structure as the previous screenshot, showing 'Administration Group' and 'Group' entries.

Display Name	Name ↑	Created	Default pinboards
A Administration Group	Administrator	8 years ago	0
G Group	Group	21 hours ago	0

5. Click **Delete**.

The screenshot shows the ThoughtSpot Admin Console with the 'ADMIN' tab selected. On the left, a sidebar lists various administrative sections: User Management, Users, Groups (which is selected and highlighted in blue), Authentication (Local, Active Directory, SAML), Settings (SSL, Reverse SSH, SMTP, Help Customization, Action Customization), and System (Overview, Cluster). The main content area displays a table of groups. At the top of the table are buttons for '+ Add Group' and 'Delete'. A search bar is also present at the top right. The table has columns for 'Display Name', 'Name ↑', 'Created', and 'Default pinboards'. Two groups are listed: 'Administration Group' (Adminstrator, created 8 years ago, 0 pinboards) and 'Group' (Group, created 21 hours ago, 0 pinboards). Both rows have a checkbox column with a red circle containing a white letter 'A' next to them.

List group members

When browsing through users or subgroups, you can often see only a limited list. To check for other users, search for the name of a specific user or subgroup.

Add multiple users to a group

To add multiple users to a group, you must be on the **Users** interface. Follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Groups** from the side navigation bar that appears.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A Administrator	tsadmin	8 years ago

4. Select the names of users you plan to add to groups by clicking the box next to the user name.

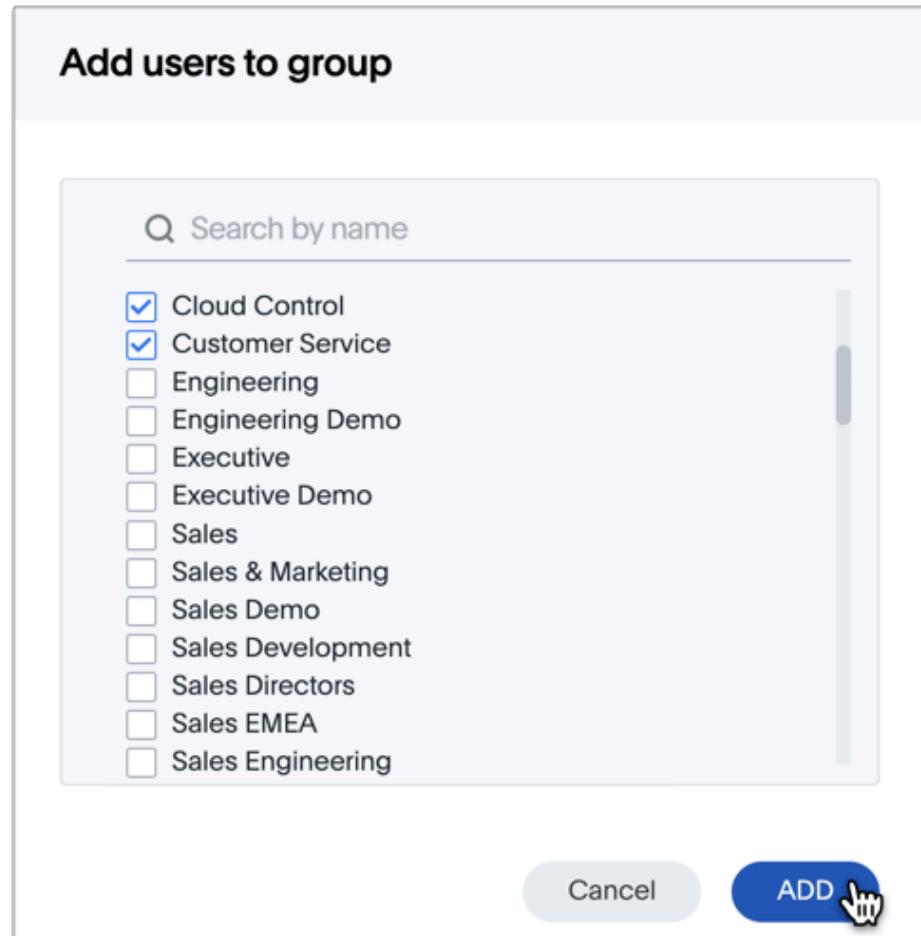
If you don't immediately see the user name, try searching for it.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A Administrator	tsadmin	8 years ago

5. Click the **Add Users to Groups** button on the top of the list of users.

Display Name	Name ↑	Created
<input checked="" type="checkbox"/> J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
<input checked="" type="checkbox"/> P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
<input type="checkbox"/> A tsadmin	tsadmin	8 years ago

6. In the **Add Users to Groups** interface, select the groups by clicking the box next to the group name.
7. Click **Add**.



Create, edit, or delete a user

Summary: For each unique person who accesses ThoughtSpot, you must create a user account. When you create a user manually in ThoughtSpot, you continue to manage that user in ThoughtSpot.

You can create users directly in ThoughtSpot, or import users and user groups through LDAP or similar protocols.

For users who have access through LDAP, the LDAP installation manages all user information.

Create a user in ThoughtSpot

This procedure demonstrates how to create a user manually. When you create a user, you can specify the *username* [See page 289], *display name* [See page 289], *sharing visibility* [See page 289], the *password* [See page 290], *email* [See page 290], whether they get a *welcome email* [See page 290] and its *text* [See page 0], and assign *group* [See page 290] memberships. The user inherits privileges and permissions directly from the group assignments.

Note that all users automatically belong to the group **All**.

To create a new user and assign that user to groups, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Users** from the side navigation bar that appears.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A tsadmin	tsadmin	8 years ago

4. Click the **+ Add User** button on the upper left-hand side of the list of users.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A tsadmin	tsadmin	8 years ago

5. In the **Add a new user** interface, enter the details for the new user:

Add a new user

Username *	Auser
Display name *	Auser
Sharing visibility *	SHARABLE
Change password *	*****
Confirm password *	*****
Email *	auser@thoughtspot.com
<input checked="" type="checkbox"/> Send a welcome email Welcome!	

* Required field

Groups

6 Groups assigned to User

Search by name

Clear all | Select all

- Manage Data
- Marketing
- Marketing Demo
- Sales Demo
- Sales Development
- Sales Directors
- Sales EMEA
- Sales Engineering
- Sales Executives

Cancel

Field	Required?	Description
Username	Yes	A login name for the user. Usernames must be unique and lowercase. If you are using Active Directory to authenticate users, and your LDAP configuration requires users to be created manually (they are not created automatically in ThoughtSpot upon authentication), the username you specify has to be domain-qualified (<code>user-name@ldap.thoughtspot.com</code>), and you must enter a dummy password.
Display name	Yes	A unique name for the user (usually their first and last name).
Sharing visibility	Yes	Indicate whether objects can be shared with this user. When set to SHARABLE , this user is an option in the Share dialog.

Change password	Yes	A password. Your password must contain three of the following: uppercase letters A-Z, lowercase letters a-z, special characters !#\$ etc, numbers 0-9. Your password must be at least eight characters long. When you click on the change password field, a tooltip appears to tell you these requirements.
Confirm password	Yes	Enter the password again.
Email	Yes	The user's email address. ThoughtSpot uses this for notification when another user shares something with them, for onboarding, for the Ask an Expert feature, and others. Note that during cluster configuration, the domain is specified. ThoughtSpot does not accept emails outside this domain. Contact ThoughtSpot Support [See page 0] to add new domains to a list of allowed domains.
Send a welcome email	No	When checked, this option ensures that the new user receives a welcome email.
Email message text	No	Enter text of the optional welcome email here.
Groups	Recommended	Select the groups for the user. Note that if you add the user to a group that has the privilege Has administration privileges , they can see all the data in ThoughtSpot. When you create a new user, the groups they belong to define the following attributes for the user: <ul style="list-style-type: none"> • Privileges: the actions they can perform, defined when you Add a group and set security privileges [See page 271]. • Permissions: the data they can access and view, defined when you consider Data security [See page 308].
		Administrators can see all data sources, and Row level security [See page 349] does not apply to them.

6. Click **Add** to create the user.

Note that this process of identifying the user's needs contributes to a robust onboarding process. See [Onboarding users \[See page 263\]](#).

Edit an existing user

As an administrator, you can edit a user account, and change the user's groups. You can also help users by resetting their password, and evaluating their onboarding experience to ensure they receive the best possible introduction to relevant information in ThoughtSpot.

To edit an existing user, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Users** from the side navigation bar that appears.

A screenshot of the ThoughtSpot Admin Console User Management screen. On the left is a vertical sidebar with a tree view of configuration categories. The "Users" node under "User Management" is currently selected and highlighted with a blue box. The main content area shows a table of users. The columns are "Display Name", "Name ↑", and "Created". There are three rows in the table:

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A Administrator	tsadmin	8 years ago

The table has a light gray background with white text. The user icons (J, P, A) are circular with colored initials. The sidebar also includes sections for Groups, Authentication (Local, Active Directory, SAML), Settings (SSL, Reverse SSH, SMTP), Help Customization, Action Customization, Backup & Storage (NAS Configuration, Snapshot), System (Overview, Cluster), and a bottom section for Overview and Cluster.

4. Click the username in the list to open the **Edit User** interface.

If you don't immediately see the username you plan to edit, try searching for it.

Display Name	Name	Created
john@thoughtspot.com	john@thoughtspot.com	21 hours ago
paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
Administrator	tsadmin	8 years ago

5. In the **Edit User** interface, edit the basic user information.

You can change the [username \[See page 289\]](#), [display name \[See page 289\]](#), [sharing visibility \[See page 289\]](#), [passwords \[See page 290\]](#), and [user's email \[See page 290\]](#).

Edit User

Username *	Auser
Display name *	Auser
Sharing visibility *	SHARABLE
Change password	
Confirm password	
Email *	auser@thoughtspot.com

Groups
6 Groups assigned to User

Search by name

Manage Data
 Marketing
 Marketing Demo

Sales Demo
 Sales Development
 Sales Directors
 Sales EMEA
 Sales Engineering
 Sales Executives

[Preview onboarding](#)

* Required field

[Cancel](#)

[Update](#)

You can also [Preview onboarding \[See page 293\]](#), and make changes to the [Groups \[See page 293\]](#) assigned to the user.

6. Click **Update**.

Preview onboarding

You can click **Preview onboarding** to evaluate this user's first experience with ThoughtSpot. After previewing the user's default data source and Pinboards, you may choose to change the **Group [See page 0]** assignments.

Edit User

Username *	Auser
Display name *	Auser
Sharing visibility *	SHARABLE
Change password	
Confirm password	
Email *	auser@thoughtspot.com

Groups **Email**
6 Groups assigned to User

Search by name

Manage Data
 Marketing
 Marketing Demo

Sales Demo
 Sales Development
 Sales Directors
 Sales EMEA
 Sales Engineering
 Sales Executives

Preview onboarding 

* Required field

Cancel Update

Groups

Follow these steps to change the user's groups:

1. Click the **Groups** tab.
2. Select the groups you want to add in the list by clicking the box next to the group name.
3. You can also use **Search** to find groups by name.

4. Deselect the groups you want to remove from the list by clearing the box next to the group name.
5. Click **Update** to save changes.

Edit User

Username *	Auser
Display name *	Auser
Sharing visibility *	SHARABLE
Change password	
Confirm password	
Email *	auser@thoughtspot.com

Groups
7 Groups assigned to User

Search by name

Manage Data
 Marketing
 Marketing Demo

Sales Demo
 Sales Development
 Sales Directors
 Sales EMEA
 Sales Engineering
 Sales Executives

Preview onboarding

* Required field

Cancel **Update** 

Delete users

To delete users, follow these steps:

1. Log in to ThoughtSpot from a browser.
2. Navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar.



3. Select **Users** from the side navigation bar that appears.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A Administrator	tsadmin	8 years ago

4. Select the users you plan to delete by clicking the box that appears when you hover over the username.

If you don't immediately see the user you plan to delete, try searching for it.

Display Name	Name ↑	Created
J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
A Administrator	tsadmin	8 years ago

5. Click **Delete**.

Create, edit, or delete a user

The screenshot shows the ThoughtSpot Admin interface with the 'User Management' section selected. On the left sidebar, under 'User Management', the 'Users' option is highlighted. The main content area displays a list of users with columns for 'Display Name', 'Name ↑', and 'Created'. Three users are listed: 'john@thoughtspot.com' (created 21 hours ago), 'paul@thoughtspot.com' (created 20 hours ago), and 'tsadmin' (created 8 years ago). The first two users have their checkboxes checked, while the third has an unchecked checkbox. A search bar at the top right contains the placeholder 'Search'.

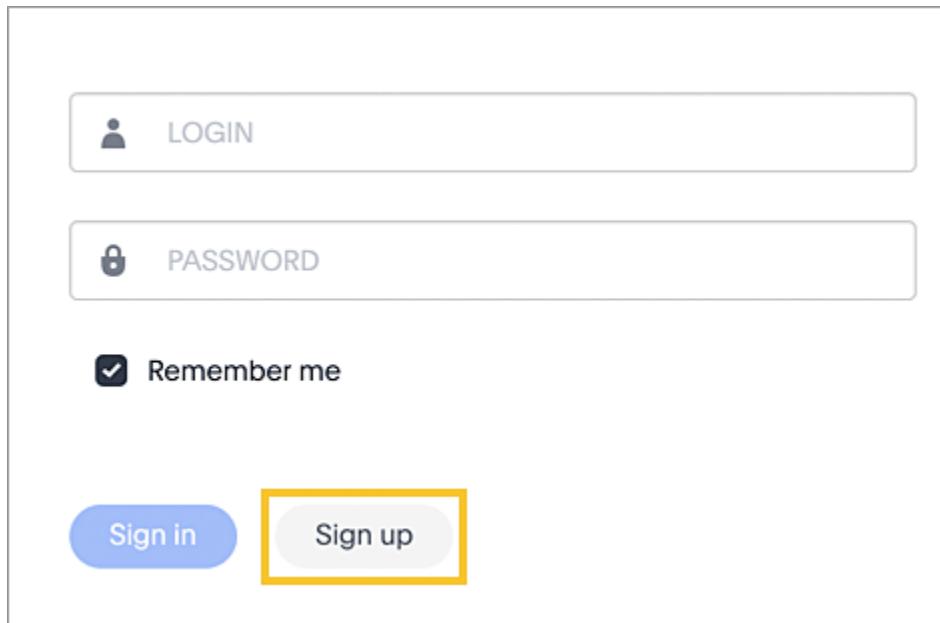
Display Name	Name ↑	Created
<input checked="" type="checkbox"/> J john@thoughtspot.com	john@thoughtspot.com	21 hours ago
<input checked="" type="checkbox"/> P paul@thoughtspot.com	paul@thoughtspot.com	20 hours ago
<input type="checkbox"/> A tsadmin	tsadmin	8 years ago

Allow users to sign up

Summary: Learn how to allow people in your organization to sign up for ThoughtSpot.

You can allow people in your organization to sign up for ThoughtSpot by clicking a button on the sign-in page.

You do this by providing them with the ThoughtSpot sign-up URL. When they go to this URL, they see the sign-up button.



When a person clicks the sign-up button, they go to a sign-up page that you've already set up outside of ThoughtSpot. This can be any page you want to use for registering new users.

Create the ThoughtSpot sign-up URL

Follow this syntax for creating the URL:

```
https://<your-thoughtspot-URL>/?signUpEnabled=true&signUpButtonLink=https://<your-sign-up-page-URL>
```

Example:

```
https://thoughtspot.mycompany.com/?signUpEnabled=true&signUpButtonLink=https://signup.mycompany.com/
thoughtspot
```

Customize the sign-up button text

By default, the button text is ‘Sign up’, but you have the option to change it. To do that, you add the ‘signUpButtonText’ parameter to the URL and include the custom text you want. If the text contains spaces, you must replace each space with a percent sign and the number 20 (%20). For example, if you wanted the button text to be ‘Request Account’, you would use ‘Request%20Account’.

Follow this syntax for creating the URL with custom button text:

```
https://<your-thoughtspot-URL>/?signUpEnabled=true&signUpButtonText=<your-sign-up-
button-text>&signUpButtonLink=https://<your-sign-up-page-URL>
```

Example:

```
https://thoughtspot.mycompany.com/?signUpEnabled=true&signUpButtonText=Request%20Account&signUpButtonLink=https://
thoughtspot
```

Display the sign-up button by default

If you want to display the sign-up button on the ThoughtSpot sign-in page without using the special sign-up URL, contact ThoughtSpot support.

Overview of security features

Summary: Learn about ThoughtSpot's security features.

There are several aspects of security, including access and permissions, data security and privacy, and security from an IT perspective.

- [System Security \[See page 300\]](#) refers to audit logs and security policies.
- [Data Security \[See page 308\]](#) refers to which users can see which data in the ThoughtSpot application, and includes:
 - [Users and Groups \[See page 265\]](#)
 - [Privileges \[See page 265\]](#)
 - [Table and columns sharing \[See page 314\]](#)
 - [Row level security \[See page 349\]](#)
 - [Worksheet sharing \[See page 318\]](#)
 - [Pinboard sharing \[See page 326\]](#)
- Network Security refers to ports for external traffic and traffic within the cluster. Some ports must remain open for handling network requests from outside the ThoughtSpot instance. To see a list of network ports that must remain open to outside traffic, and for inter-cluster communication, review the information in [Network ports \[See page 148\]](#).

System security tools and processes

Summary: System security refers to audit logs and security policies.

ThoughtSpot includes a number of management tools, monitoring applications, and automated processes to support system security. System security includes managing access and privileges, audit logs, security policies, and Linux OS installed package updates.

Audit logs

There are several ways you can view audit log information in ThoughtSpot. You can see recent events in the Control Center or view more detailed audit logs using tscli. Administrators can view audit logs of configuration changes users have made to ThoughtSpot in these ways:

- Monitor events from the [Control Center \[See page 613\]](#).
- Generate audit log reports through the `tscli` command.

You can access an audit log of cluster events through tscli. You can also access information on cluster updates, configurations, data loading and metadata events.

Use the `tscli event list` command to return an audit list of events from the cluster. The syntax is:

```
tscli event list
[--include <all|config|notification>]
[--since <hours,minutes,days>
 | --from <yyyymmdd-HH:MM>
 --to <yyyymmdd-HH:MM>]
[--detail]
[--summary_contains
<'string1'| 'string2' ...>]
[--detail_contains
<'string1'| 'string2' ...>]
[--attributes
<key1='value1'| 
key2='value2' ...>]
```

Optional parameters are:

Parameter	Description
--include	Specifies the type of events to include, and can be <code>all</code> , <code>config</code> , or <code>notification</code> .
--detail	Returns the events in a detail format rather than a tabular summary, which is the default.
--summary_contains <'string1' 'string2' ...>	Specifies a string to check for in the event summary. Enclose strings in single quotes, and separate multiple strings with <code> </code> . Events that match all specified strings will be returned.
--detail_contains <'string1' 'string2' ...>	Specifies a string to check for in the detail. Enclose strings in single quotes, and separate multiple strings with <code> </code> (pipe symbol). Events that match all specified strings will be returned.
--attributes <key1='value1' &pipe; key2='value2' ...>	Specifies attributes to match as key=value pairs. Separate multiple attributes with <code> </code> (pipe symbol). Events that match all specified key/value pairs will be returned. Put single quotes around the value(s).

And a time window made up of either:

- `--since <hours,minutes,days>` is a time in the past for where the event audit begins, ending at the present time. Specify a human readable duration string, e.g. `4h` (4 hours), `30m` (30 minutes), `1d` (1 day).

Or both:

- `--from <yyyymmdd-HH:MM>` is a timestamp for where to begin the event audit. It must be of the form: `yyyymmdd-HH:MM`.
- `--to <yyyymmdd-HH:MM>` is a timestamp for where to end the event audit. It must be of the form: `yyyymmdd-HH:MM`.

To get audit logs:

1. Log in to the Linux shell using SSH.
2. Issue the `tscli event list` command, with the desired parameters, for example:

```
$ tscli event list
  --include config
  --since 24 hours
```

Security policies

Security policies are the principles and processes ThoughtSpot uses in development to ensure a product that conforms to security standards. Security policies ensure a secure product with each release. When a release is in development, each build is tested using Qualys Network Security and Vulnerability Management Suite. Issues and vulnerabilities are fixed proactively, based on the results.

The ThoughtSpot Engineering and ThoughtSpot Support teams are notified of Common Vulnerabilities and Exposures (CVEs), so they can patch OS packages proactively as well. You can view installed packages along with their version numbers at any time, in order to see if you require an update to ThoughtSpot.

Whenever a CVE is identified, and an OS package needs to be updated, the next patch release will include the patch or update. You can view installed Linux packages at any time, along with the version numbers of the installed packages.

Third-party security software for security, governance, and monitoring of ThoughtSpot

You can install supported [third-party security and monitoring software \[See page 303\]](#) on a ThoughtSpot cluster.

About third-party security and monitoring software

Summary: You can install third-party software for security, governance, and monitoring of ThoughtSpot.

In addition to the ThoughtSpot monitoring and security features, some companies require specific additional third-party software to comply with their internal IT policies. This allows them to support all of their systems with a common set of security and management tools.

For example, you may wish to accomplish some security and monitoring tasks with your own third-party software. These tasks include things like pushing alerts, events, forensics, audit trails, insights, etc. from ThoughtSpot to your own local monitoring systems.

Supported third-party software

ThoughtSpot supports installation of the following third-party software on the ThoughtSpot instance:

- Qualys Qualys is a widely used technical vulnerabilities and security compliance scanning tool. For more information about Qualys, see the [Qualys documentation](http://www.qualys.com/documentation/) (<http://www.qualys.com/documentation/>).
- SNMP (Simple Network Management Protocol) SNMP is an industry standard protocol used for monitoring network traffic and alert events.
- Splunk You can install Splunk rsyslog and use it to forward ThoughtSpot logs to Splunk. For more information about Splunk, see the [Splunk documentation](http://docs.splunk.com/) (<http://docs.splunk.com/>).

Install third-party software

For details on how to install third-party software, see: [Installing third-party security and monitoring software \[See page 306\]](#)

What is not supported

When installing and configuring third-party software on a ThoughtSpot cluster, follow the following guidelines to avoid interfering with cluster operations:

- Avoid making any direct changes to any files outside of the /home directory.
- Do not remove existing SSH keys or authorized keys from /home/admin/.ssh
- Excessive resource usage, e.g. CPU, disk, memory, processes, etc.
- Killing any system or ThoughtSpot services, or causing node reboots.

Do not change any system wide configuration which may affect ThoughtSpot, such as:

- Network, e.g. IP addresses, DNS resolution
- Storage, e.g. removing existing mount points, removing drives
- Security, e.g. selinux

Qualys

Qualys is supported for scanning of ThoughtSpot clusters for security vulnerabilities.

SNMP Traps

ThoughtSpot has a built-in alerting service that can also be used to send SNMP traps. Many third-party monitoring systems share the common standard of using SNMP traps, and you can take advantage of those capabilities with ThoughtSpot.

ThoughtSpot supports SNMP for read only. So for example, you can read the IP address of the cluster, but not change it using SNMP.

See the [Alert code reference \[See page 0\]](#) for details.

Splunk rsyslog

ThoughtSpot monitoring and alerting logs are written to standard locations in the file system. This allows you to use rsyslog to collect them and send them to Splunk.

Here are some links to help you learn where various logs are written in ThoughtSpot:

- [Monitoring logs \[See page 587\]](#)
- [Audit logs \[See page 300\]](#)
- [Alert code reference \[See page 0\]](#)

Installing third-party security and monitoring software

Summary: You can install third-party software for security, governance, and monitoring of ThoughtSpot.

This procedure shows how to install supported [third-party security and monitoring software](#) [See page 303] on a ThoughtSpot cluster:

To install third-party software

1. Log in to the Linux shell using SSH.
2. Issue the `tscli ansible checkout` command, specifying a temporary directory, for example:

```
$ tscli ansible checkout  
Checking out playbooks successfully in /tmp/111895937.
```

3. Switch to the temporary directory that was created.

```
$ cd /tmp/111895937
```

4. In the temporary directory, save or edit the playbooks and modules.

If you want to create a global ordering between playbooks, name them in alphabetical order, e.g. 10.first.yml, 20.second.yml, etc. You can also specify the order line by line in order.txt within the same directory.

5. Commit your changes. This command will validate the playbook first, and then apply it.

Use the `-local` flag if you want to commit the change only to local storage on the local node. Otherwise, push it will go to centralized storage, and your changes will apply to all nodes in the cluster.

- To apply your changes globally to all nodes in the cluster, issue the command:

```
$ tscli ansible commit
```

- To apply your changes on the local node only, issue the command:

```
$ tscli ansible commit --local
```

This commits your changes. If there is a problem with the configuration, you will see an error message in standard output.

Data security

Summary: Data security refers to which users can see which data in the ThoughtSpot application.

Sharing and security privileges govern what data a user can access and what they can do with the data. Admins can use privileges to regulate access to information and provide a personalized user experience.

Users, groups, and privileges

Data security applies to users and groups. Users can be managed [manually \[See page 265\]](#) or through [LDAP \[See page 117\]](#). Each user can have membership in one or more groups. Admins can make security settings that determine what users are allowed to do in ThoughtSpot. These settings are applied at the group level.

The following table shows the intersection of user privilege and ability:

	Create/Edit WS	Create View	Create Embrace Connection	Modify Col. Props. ¹	Upload Data	Download Data	Share within Group	Share with all users	Manage RLS rules	CrUD Relationships	Read Relationships	See Hidden Cols	Join with Upload Data	Schema Viewer	Use Scheduler	Use Auto-Analyze
Can administer ThoughtSpot	Y	Y	Y	Y	Y	Y	Y	Y	Y ²	Y	Y	Y	Y	Y	Y	Y
Can upload user data	N	N	N	N	Y	Y	Y	Y	Y ³	Y ⁴	N	N	N	N	N	N
Can download data	N	N	N	N	N	Y	Y	N	N	Y ⁴	N	N	N	N	N	N
Can manage data	Y	Y	Y	Y	Y	Y	N	Y	Y ⁴	Y ⁴	Y ⁵	Y	N	N	N	N
Can share with all users	N	N	N	N	N	Y	Y	N	N	Y ⁴	N	N	N	N	N	N
Has SpotIQ privilege	N	N	N	N	N	N	N	N	N	Y ⁴	N	N	N	N	N	Y
Can Administer and Bypass RLS	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
None	N	N	N	N	Y	N	N	Y ⁴	N	N	N	N	N	N	N	N

Table notes:

1. Applies to non-owners only.
2. Any tables.
3. Author of at least one table in relationship.
4. Only when read permission for columns used in the relationship.
5. With edit permission.



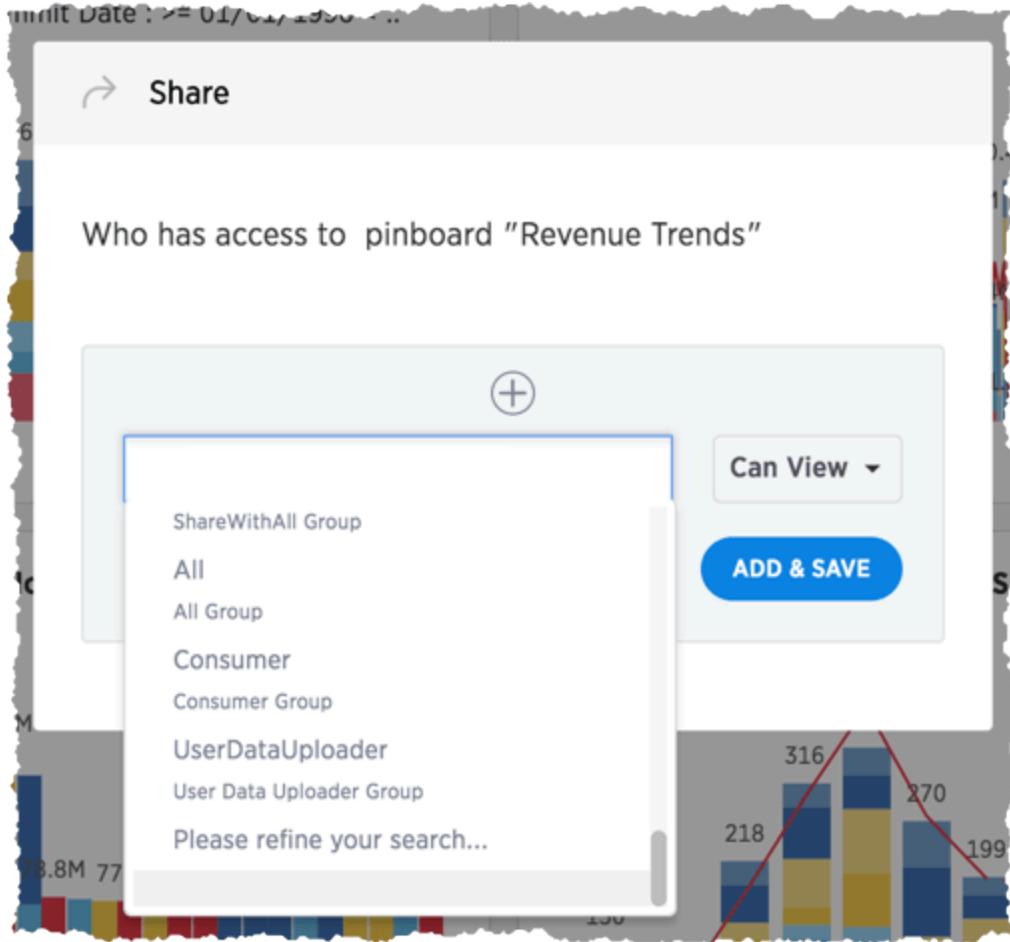
Security model for sharing objects

You can share with groups and with individual users. Sharing of tables can be defined at the table, column, or row level. This provides flexibility in modeling your data security policy. Security and sharing settings apply to several different types of objects, each of which has its own security default settings and rules.

Object type	Description	Default security model
Tables	The source data tables that have been loaded using ThoughtSpot Loader.	Administrator users have access to source tables. They can share a table with other users or groups. See Sharing tables and columns [See page 314] .
Columns	The columns in the source data tables that have been loaded using ThoughtSpot Loader.	Administrator users have access to columns in the source tables. They can share selected columns with other users or groups. See Sharing tables and columns [See page 314] .
Rows	The rows in the source data tables that have been loaded using ThoughtSpot Loader.	All rows in the source tables are shared with all users by default.
Imported data	Data that was imported using a Web browser.	Only the user who imported the data (and any user with administrator privileges) has access to it by default. They can share a table (or selected columns) with other users or groups. See Sharing tables and columns [See page 314] .
Worksheets	A worksheet created using a Web browser.	Only the creator of the worksheet (and any user with administrator privileges) has access to it by default. They can share a worksheet with other users or groups. See Share Worksheets [See page 318] .
Pinboards	A pinboard of saved search results.	Anyone who can view a pinboard can share it. See Share Pinboards [See page 326] .

Understanding SHARABLE

When you share an object, only the users and groups that have **SHARABLE** set for the **Sharing visibility** option appear on the dialog.



Only users in the **Administrators** group or users with **Admin** privileges can share with groups marked as **NOT SHARABLE**. Members of a group with **Can share with all users** authorization can also share with groups marked as **NOT SHARABLE**.

Users in groups marked **NOT SHARABLE** cannot share objects among themselves. In multi-tenant scenarios, admins can create groups that bring together portions of two non-share groups so that they can share. For example, the members of group C can share even if they belong to other groups that cannot.

Row level security

ThoughtSpot includes robust row level security, which allows you to filter all objects users see based on conditions you set at the level of row values in base data tables.

You may find it useful to create groups for RLS. To prevent these groups from appearing in the **Share** dialog, create a **NOT SHARABLE** group with a single user and an RLS group with another single user (1-to-1).

Related information

- [Revoke access \(unshare\) \[See page 346\]](#)
- [Row level security \[See page 349\]](#)

Sharing tables and columns

Summary: As an administrator, you can share view or edit access to any table.

By default, when a table is loaded using the ThoughtSpot Loader, ODBC, or JDBC, it is only visible to administrators. Tables imported from a Web browser are visible to administrators and the user who uploaded it. Administrators and owners can share **Can View** or **Can Edit** privileges on tables with other users, who can further share them with others.

When you share a table, you can share the entire table, or specific columns. For column level security, share only specific columns.

Permissive or strict sharing

Use caution when sharing tables, because any objects created from that table will have dependencies on it and its underlying structure. Objects created from tables can include worksheets, Views, Answers, and Pinboards. This means that if a user wants to drop or modify a table, any object that depends upon it must be edited or removed first, to remove the dependency.

For this reason, it is a best practice to only grant the **Edit** permission on a table to a small number of users. If you want to prevent shares from also revealing the columns regardless of where it appears (worksheets, answers, and pinboards), you can ask [ThoughtSpot Support \[See page 0\]](#) to enable a stricter behavior.

You can share a table [from the Data tab \[See page 314\]](#), or [from within the table \[See page 316\]](#) that you want to share.

Share from the Data tab

To share a table, worksheet, or View from the **Data** tab, follow these steps.

1. Click **Data** on the top navigation bar.

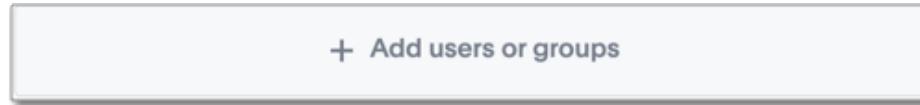
2. Hover over the dataset you want to share and select it by clicking the empty checkbox that appears.
3. Click **Share**.

4. If you are sharing a table, select **Entire Table** or **Specific Columns**.

5. If you select **Specific Columns**, select the column(s) you want to share. If you select **Entire Table**, skip to step 6.

Note: You cannot click multiple columns at once. You must input the users or groups with whom you want to share for each column.

6. Click **+ Add users or groups** and select the users and groups with whom you want to share.



7. Specify permissions [See page 317].

Share from within the dataset

To share a table, worksheet, or View from within the dataset, follow these steps.

1. Navigate to the dataset you want to share.
2. Click the ellipsis icon  , and then click **Share**.

The screenshot shows the 'MarketSpot_Vendor_Dimension' table in the Data tab. The 'Share' button in the top right corner is highlighted with a mouse cursor. The table has columns: COLUMN NAME, DESCRIPTION, DATA TYPE, COLUMN TYPE, ADDITIVE, AGGREGATION, HIDDEN, SYNONYM, and a Share icon. There are five rows of data. The bottom left corner of the table area says '5 rows total'.

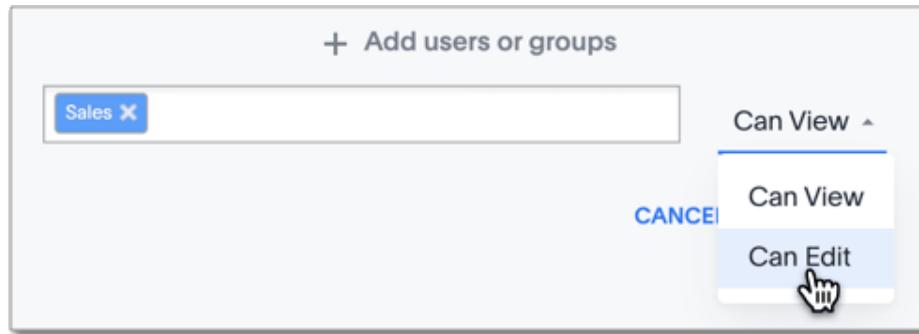
3. Select the users or groups with whom you want to share.

You can only share the entire dataset. You cannot share individual columns. To share individual columns, share a table [from the Data tab \[See page 0\]](#). You cannot share individual columns for worksheets or Views.

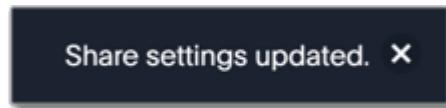
4. Specify permissions [See page 317].

Specify permissions

1. Configure the level of access by selecting from the dropdown list. You can select:
 - **Can View** to provide read-only access. This enables viewing the data source's data, as well as defining worksheets based on that data.
 - **Can Edit** to allow modification. This enables renaming, modifying, or deleting the entire data source and adding or removing its columns.



2. Click **Add**.
3. Click **Done**.
4. The **Share settings updated** notification appears on the bottom of your screen.



Share worksheets

Summary: You can share worksheets with users or with groups.

Sharing a worksheet allows users to select it as a data source and search it.

When you share a worksheet, you give users or groups within your cluster access to the worksheet's data. You share all of its columns. Sharing a worksheet does not share the underlying tables. If you want to share the underlying tables, see [Sharing tables and columns \[See page 314\]](#).

If you want to migrate the worksheet to another cluster, or otherwise export it, see [Migrate or restore objects \[See page 521\]](#).

A worksheet can be shared by the owner of the worksheet, or by an administrator. Users can start searching a worksheet as soon as the worksheet is shared with them.

You can share a worksheet from the list of worksheets under **worksheets** on the **Data** tab [See page 318], or from [within the worksheet \[See page 320\]](#).

Share from the Data tab

To share a table, worksheet, or View from the **Data** tab, follow these steps.

1. Click **Data** on the top navigation bar.
2. Hover over the dataset you want to share and select it by clicking the empty checkbox that appears.
3. Click **Share**.

The screenshot shows a list of tables in the ThoughtSpot interface. The 'Share' button is highlighted for the 'MarketSpot_Vendor_Dimension' table. The table details are as follows:

Name	Source	Stickers	Materialize Status	Modified	Author
TRIPS_VW	MarketSpot		3 months ago	E	Engineering Admin
MarketSpot_Vendor_Dimension	MarketSpot		9 months ago	A	Administrator Super-User
MarketSpot_Wholesale_Orders_Fact	MarketSpot		9 months ago	A	Administrator Super-User
MarketSpot_Department_Dimension	MarketSpot		9 months ago	A	Administrator Super-User
MarketSpot_Sales_Fact	MarketSpot		9 months ago	A	Administrator Super-User
MarketSpot_Product_Dimension	MarketSpot		9 months ago	A	Administrator Super-User
MarketSpot_Store_Dimension	MarketSpot		9 months ago	A	Administrator Super-User

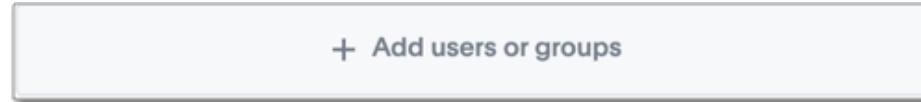
- If you are sharing a table, select **Entire Table** or **Specific Columns**.

The screenshot shows the 'Share' dialog for the 'MarketSpot_Vendor_Dimension' table. The 'Specific Columns' option is selected, and the 'Vendor_Key' column is highlighted. The dialog also shows other columns: Vendor_Nbr, Purchase_Manager, Manager_Start_Date, and Manager_End_Date.

- If you select **Specific Columns**, select the column(s) you want to share. If you select **Entire Table**, skip to step 6.

Note: You cannot click multiple columns at once. You must input the users or groups with whom you want to share for each column.

- Click **+ Add users or groups** and select the users and groups with whom you want to share.



7. Specify permissions [See page 320].

Share from within the dataset

To share a table, worksheet, or View from within the dataset, follow these steps.

1. Navigate to the dataset you want to share.
2. Click the ellipsis icon  , and then click **Share**.

A screenshot of the ThoughtSpot Data tab interface. At the top, there are several tabs: SEARCH, ANSWERS, PINBOARDS, SPOTIQ, DATA (which is selected), and ADMIN. Below the tabs, the title 'SYSTEM TABLE MarketSpot_Vendor_Dimension' is displayed. Underneath the title is a table with columns: COLUMN NAME, DESCRIPTION, DATA TYPE, COLUMN TYPE, ADDITIVE, AGGREGATION, HIDDEN, and SYNONYM. There are five rows of data in the table. To the right of the table, there is a sidebar with options: Review Suggestions (0), SpotIQ analyze, Manage experts, and a prominent 'Share' button with a hand cursor icon over it. At the bottom left of the table area, it says '5 rows total'.

3. Select the users or groups with whom you want to share.

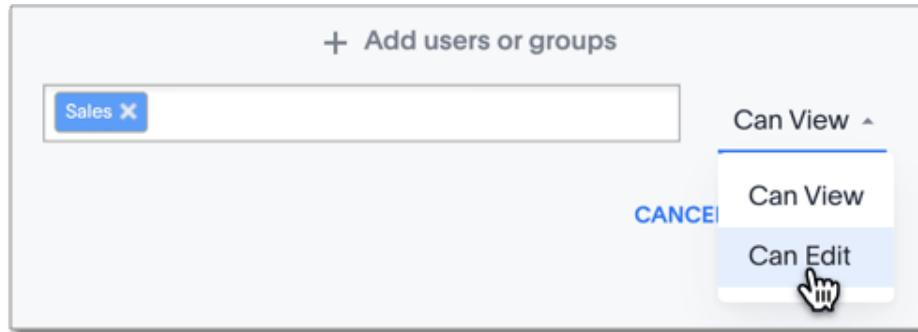
You can only share the entire dataset. You cannot share individual columns. To share individual columns, share a table [from the Data tab \[See page 0\]](#). You cannot share individual columns for worksheets or Views.

4. Specify permissions [See page 320].

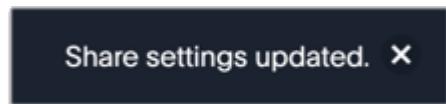
Specify permissions

1. Configure the level of access by selecting from the dropdown list. You can select:

- **Can View** to provide read-only access. This enables viewing the data source's data, as well as defining worksheets based on that data.
- **Can Edit** to allow modification. This enables renaming, modifying, or deleting the entire data source and adding or removing its columns.



2. Click **Add**.
3. Click **Done**.
4. The **Share settings updated** notification appears on the bottom of your screen.



Share Views

Summary: You can share Views with users or with groups. Sharing a View allows users to select it as a data source and search it.

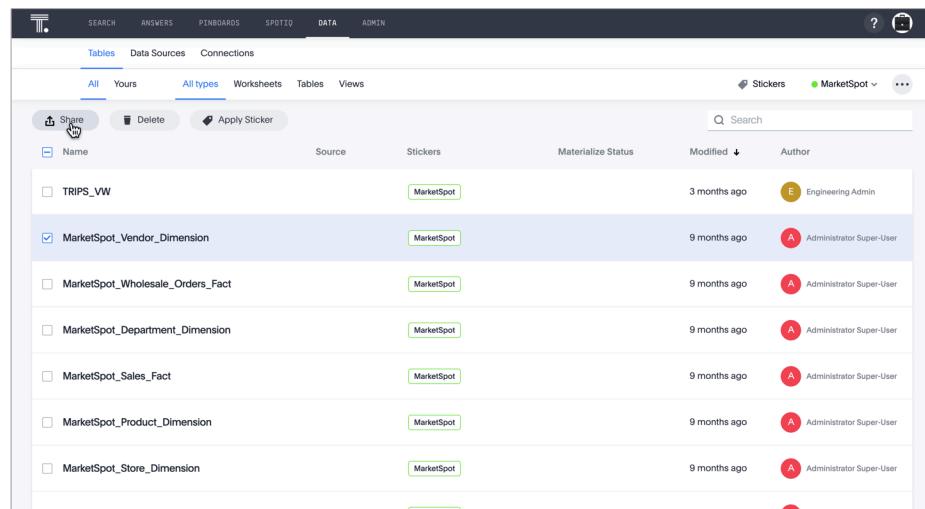
When you share a View, you share all of its data. Sharing a View does not share the underlying tables. If you want to share the underlying tables, see [Sharing tables and columns \[See page 314\]](#). A View can be shared by the owner of the View, or by an administrator. Users can start searching a View as soon as the View is shared with them.

You can share a View from the list of Views under **Views** on the **Data** tab [See page 322], or from the [View itself \[See page 323\]](#).

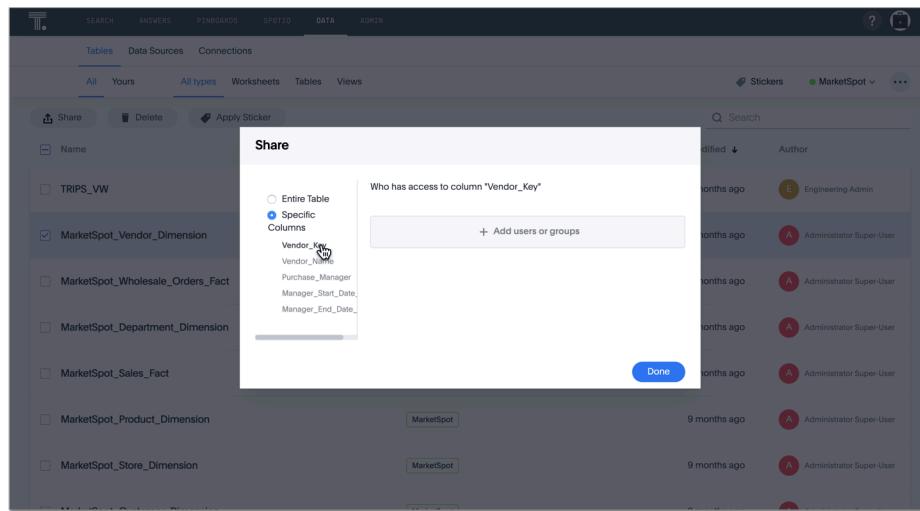
Share from the Data tab

To share a table, worksheet, or View from the **Data** tab, follow these steps.

1. Click **Data** on the top navigation bar.
2. Hover over the dataset you want to share and select it by clicking the empty checkbox that appears.
3. Click **Share**.



- If you are sharing a table, select **Entire Table or Specific Columns**.



- If you select **Specific Columns**, select the column(s) you want to share. If you select **Entire Table**, skip to step 6.

Note: You cannot click multiple columns at once. You must input the users or groups with whom you want to share for each column.

- Click **+ Add users or groups** and select the users and groups with whom you want to share.

+ Add users or groups

- Specify permissions [See page 324].

Share from within the dataset

To share a table, worksheet, or View from within the dataset, follow these steps.

- Navigate to the dataset you want to share.
- Click the ellipsis icon  , and then click **Share**.

The screenshot shows the ThoughtSpot interface with the 'DATA' tab selected. A table named 'MarketSpot_Vendor_Dimension' is displayed. The 'Share' button in the top right corner of the table view is highlighted with a mouse cursor. The table has columns for COLUMN NAME, DESCRIPTION, DATA-TYPE, COLUMN TYPE, ADDITIVE, AGGREGATION, HIDDEN, SYNONYM, and a Share icon.

COLUMN NAME	DESCRIPTION	DATA-TYPE	COLUMN TYPE	ADDITIVE	AGGREGATION	HIDDEN	SYNONYM	Share
Vendor_Key	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Vendor_Name	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Purchase_Manager	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Manager_Start_Dat...	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Manager_End_Date...	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT

5 rows total

- Select the users or groups with whom you want to share.

You can only share the entire dataset. You cannot share individual columns. To share individual columns, share a table [from the Data tab \[See page 0\]](#). You cannot share individual columns for worksheets or Views.

- Specify permissions [\[See page 324\]](#).

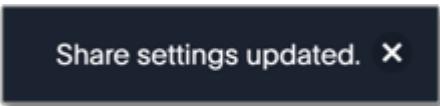
Specify permissions

- Configure the level of access by selecting from the dropdown list. You can select:

- Can View** to provide read-only access. This enables viewing the data source's data, as well as defining worksheets based on that data.
- Can Edit** to allow modification. This enables renaming, modifying, or deleting the entire data source and adding or removing its columns.

The screenshot shows a modal dialog box titled '+ Add users or groups'. It contains a text input field with 'Sales' and a 'Cancel' button. To the right, a dropdown menu is open with two options: 'Can View' and 'Can Edit'. The 'Can Edit' option is highlighted with a blue background and a hand cursor icon.

2. Click **Add**.
3. Click **Done**.
4. The **Share settings updated** notification appears on the bottom of your screen.

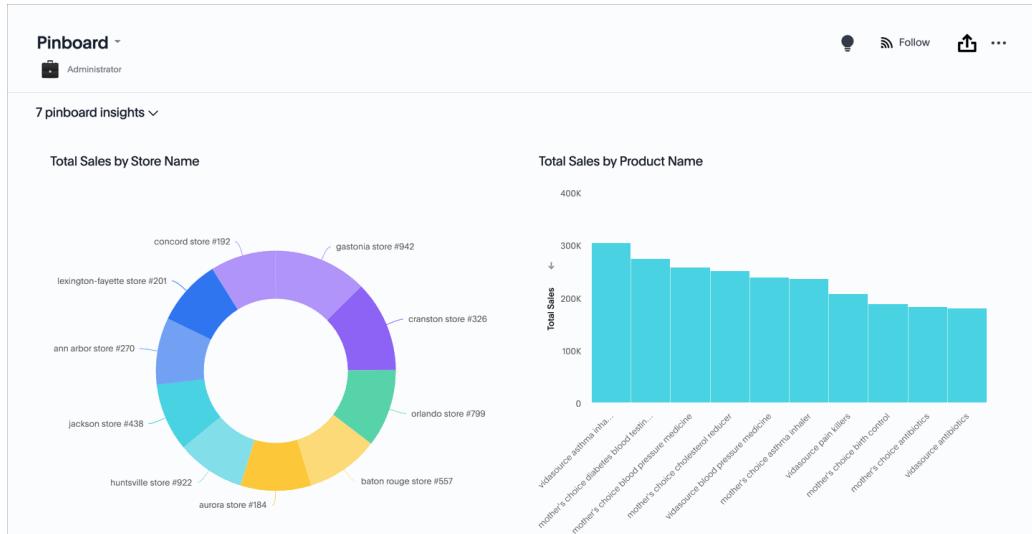


Share settings updated. X

A dark blue rectangular box with rounded corners containing the text "Share settings updated." in white. To the right of the text is a small white "X" icon.

Share a pinboard

Summary: Whenever you view a pinboard you have the option of sharing it with others.



When you share a Pinboard, you share a live link to the Pinboard that reflects the latest version of it.

When someone else views the Pinboard you shared with them, they see the most recently saved version with the most recent data.

You do not have to be an administrator or the Pinboard's owner to share saved Pinboards. Any user can share them, based on the access levels the user has.

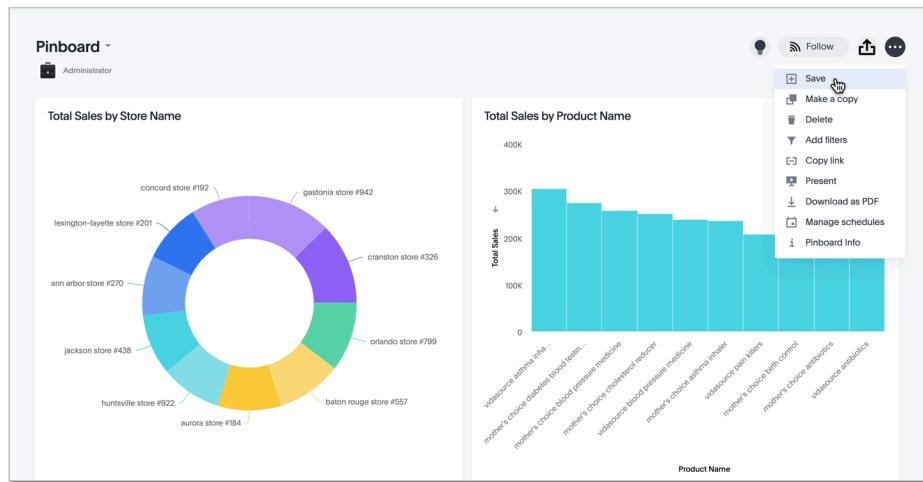
You can share a Pinboard from the list of Pinboards on [the main pinboards page \[See page 326\]](#), or from [the pinboard itself \[See page 328\]](#). You can also share [a specific visualization \[See page 328\]](#) within a Pinboard.

Share from the Pinboards page

To share a Pinboard from the main Pinboard page, follow these steps.

1. Configure the Pinboard to look as it must appear when you share it.

2. Save the Pinboard by clicking the ellipsis icon , and selecting **Save**.

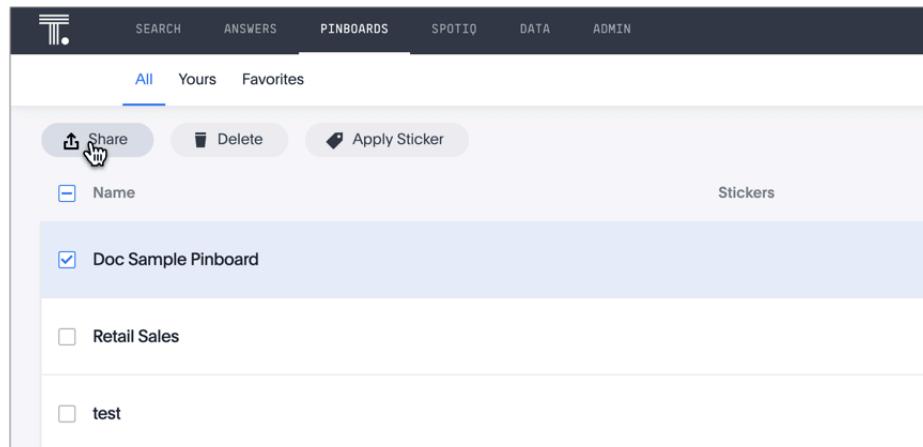


3. Click **Pinboards** on the bar at the top of your screen.



4. Select the Pinboard you want to share from the list of Pinboards by hovering over it and clicking the empty check box that appears.

5. Click **Share**.



6. Specify permissions [See page 329].

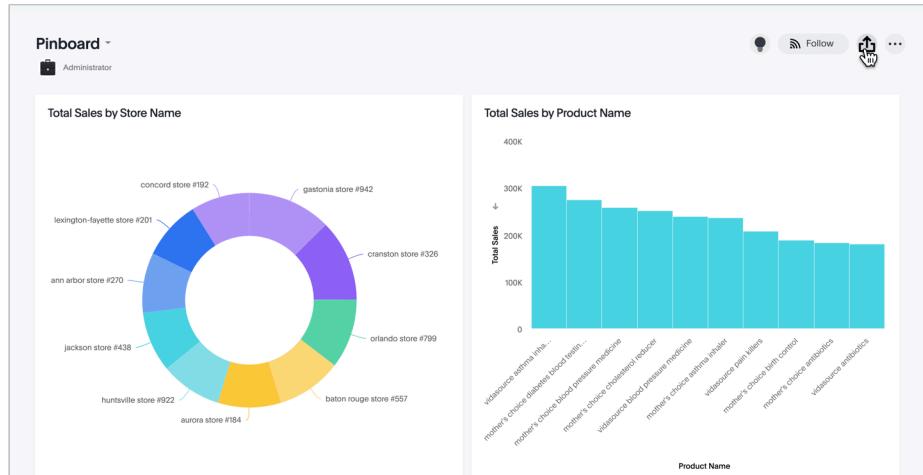
Share from within a pinboard

To share a Pinboard from within the Pinboard, follow these steps.

1. Configure the Pinboard to look as it must appear when you share it.

2. Save the Pinboard by clicking the ellipsis icon  , and selecting **Save**.

3. Click the sharing icon .



4. Specify permissions [See page 329].

Share a Pinboard visualization

You can also share a specific visualization within a Pinboard. When you share a Pinboard visualization, the user or group receives an email with a link to that visualization.

Note that sharing a specific visualization within a Pinboard gives users and groups access to the entire Pinboard.

To share a Pinboard visualization, follow these steps:

1. Navigate to the Pinboard visualization that you would like to share.

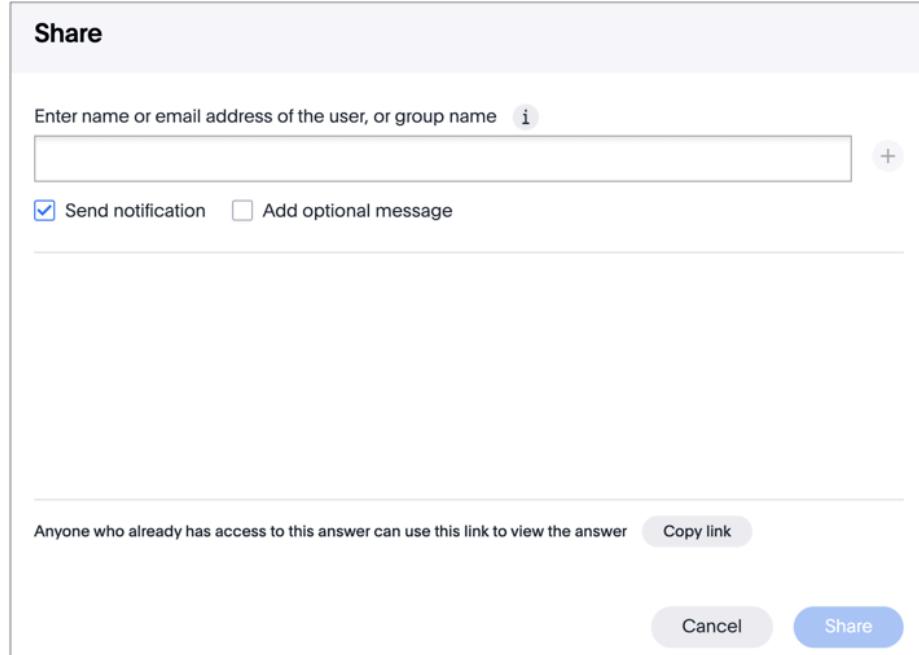
2. Select the sharing icon  that appears when you hover over the visualization.



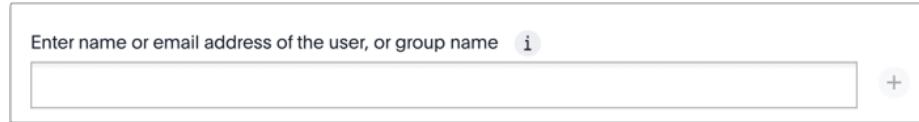
3. Specify permissions [See page 329].

Specify permissions

1. After you click the **Share** icon, the sharing dialog box appears.



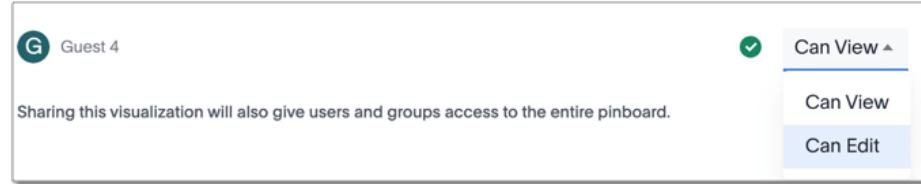
2. Enter users or groups with whom you want to share this object in the text box.



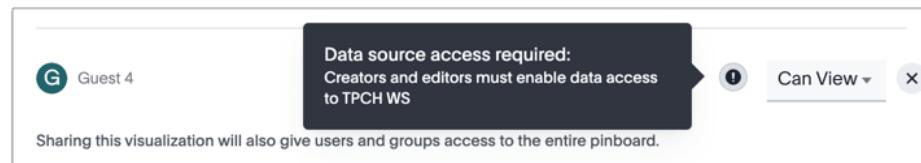
Note that you can only enter email addresses whose domains are in your list of allowed domains. These domains appear when you click on the info button **i**.

Tip: If you want to hide the allowed email domains for your company, or otherwise customize them, contact ThoughtSpot support.

3. Configure the level of access by selecting from the drop-down list. You can select:
 - **Can View** to provide read-only access. If the user doesn't have access to the underlying worksheet, they can only view the shared object.
 - **Can Edit** to allow modification. Enables renaming or deleting the shared object. If a user with edit privileges modifies a shared object, the object saves their changes.



4. If the selected group or user does not have access to the underlying data, you must enable access to the worksheet, view, or table. A black warning symbol appears when you try to share with a user who does not have underlying data access. If you click on it, it tells you to enable access:



If you own the underlying data source, you can enable access through the sharing dialog box. If you do not own the data source, ThoughtSpot emails the owner of the data source or your ThoughtSpot administrator to ask them to share the data.

To enable access, select **Give view access to underlying data sources** at the bottom of the dialog box. You can click on the arrow to view the data sources.

Share

Enter name or email address of the user, or group name [i](#)

Send notification Add optional message

G Guest 4 ✓ Can View ▾

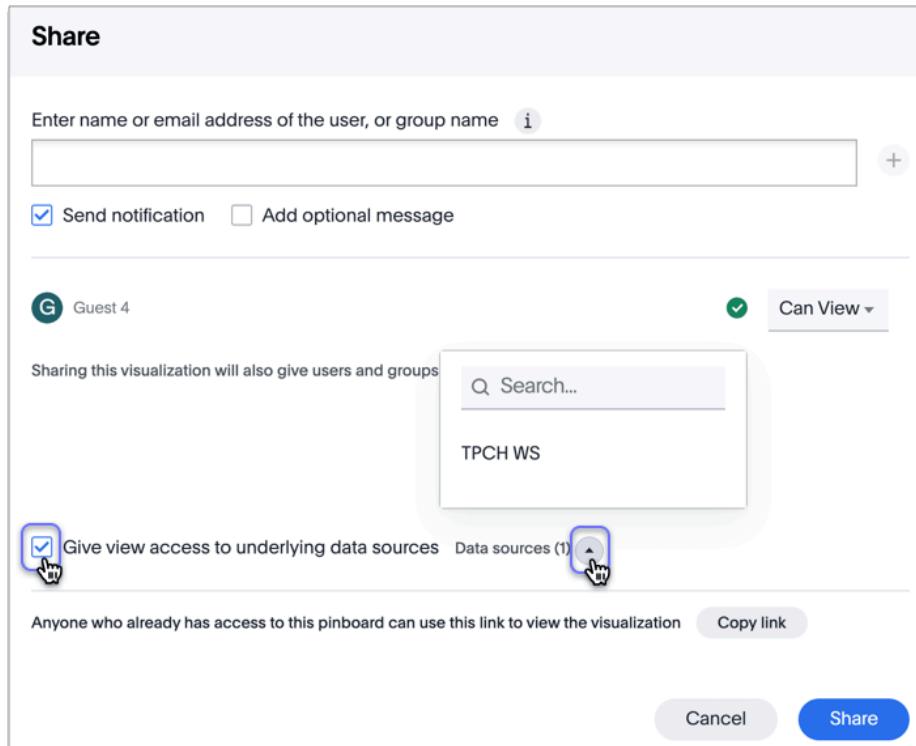
Sharing this visualization will also give users and groups:

Search...
TPCH WS

Give view access to underlying data sources Data sources (1) 

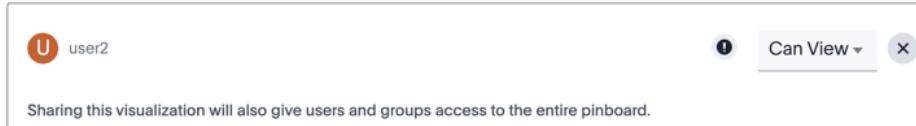
Anyone who already has access to this pinboard can use this link to view the visualization [Copy link](#)

[Cancel](#) [Share](#)

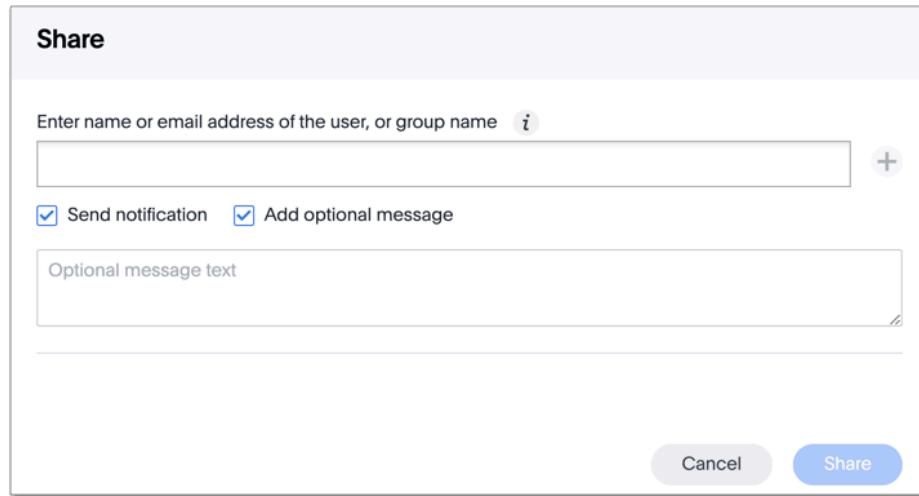


After you enable access, the warning symbol turns into a green checkmark.

5. To stop sharing with a user or group, click the **x** icon .

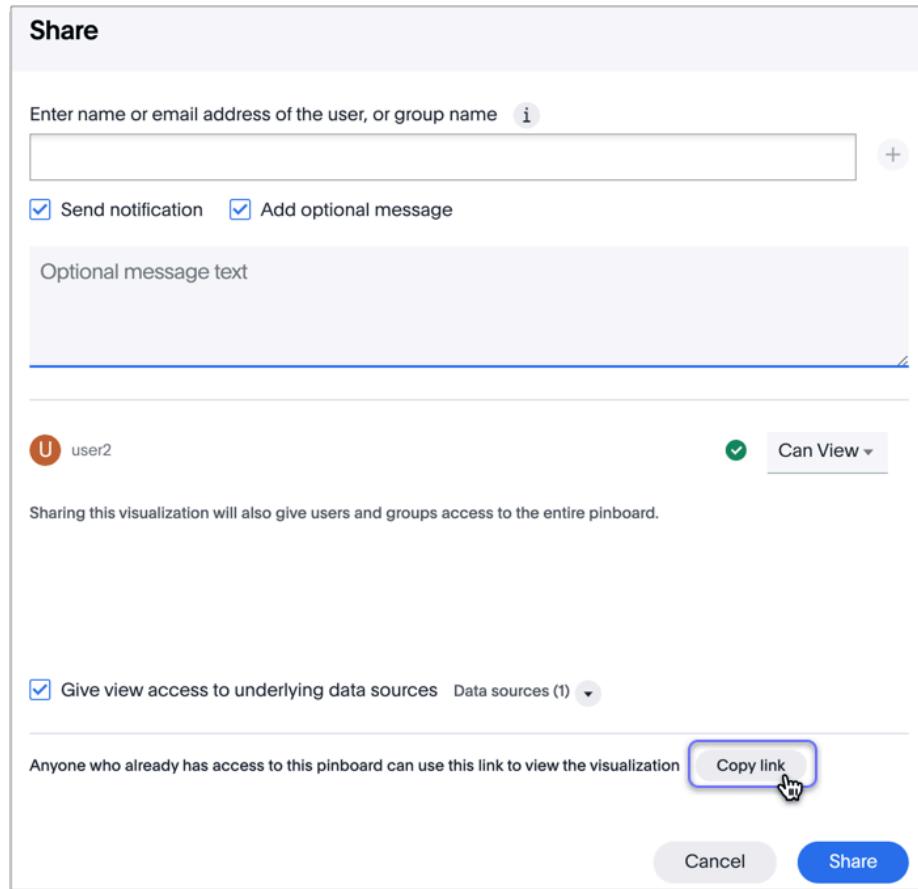


6. You can send an email notification and an optional message:



7. You can also copy a direct link to the Answer, Pinboard, or visualization within a Pinboard that you are sharing, and separately send that link to users after you share the object with them. Simply click the **Copy link** button at the bottom of the sharing dialog box.

Note that sending users this link does not share the object with them. You must also share the object by clicking the **Share** button at the bottom of the dialog box.



8. Click **Share**.

Share

Enter name or email address of the user, or group name [i](#)

Send notification Add optional message

Optional message text

 user2  Can View ▾

Sharing this visualization will also give users and groups access to the entire pinboard.

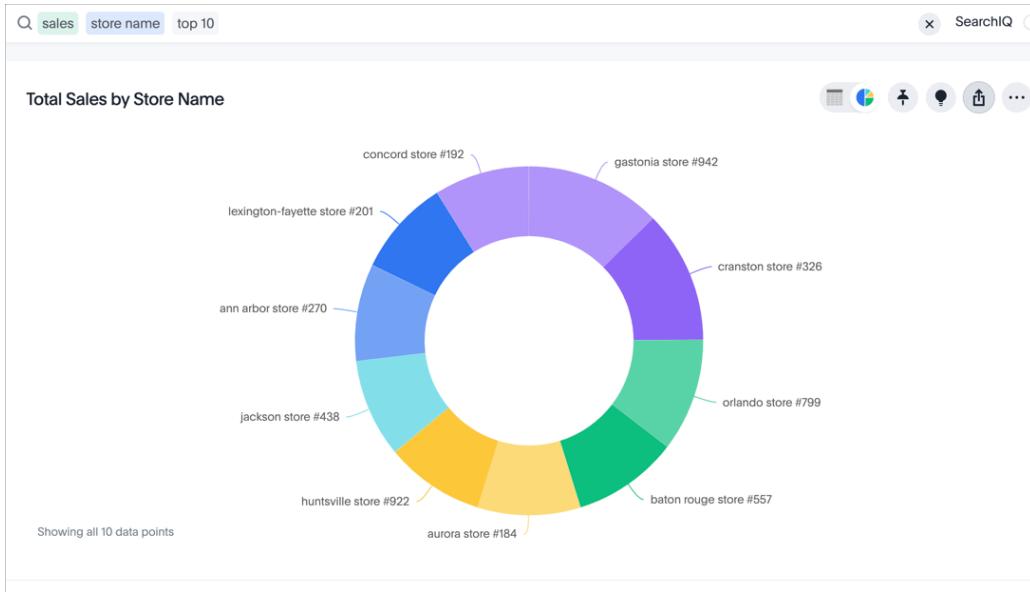
Give view access to underlying data sources [Data sources \(1\)](#)

Anyone who already has access to this pinboard can use this link to view the visualization [Copy link](#)

[Cancel](#) **Share** 

Share answers

Summary: You do not have to be an administrator or the owner to share saved answers. Any user can share them, based on the access levels the user has.



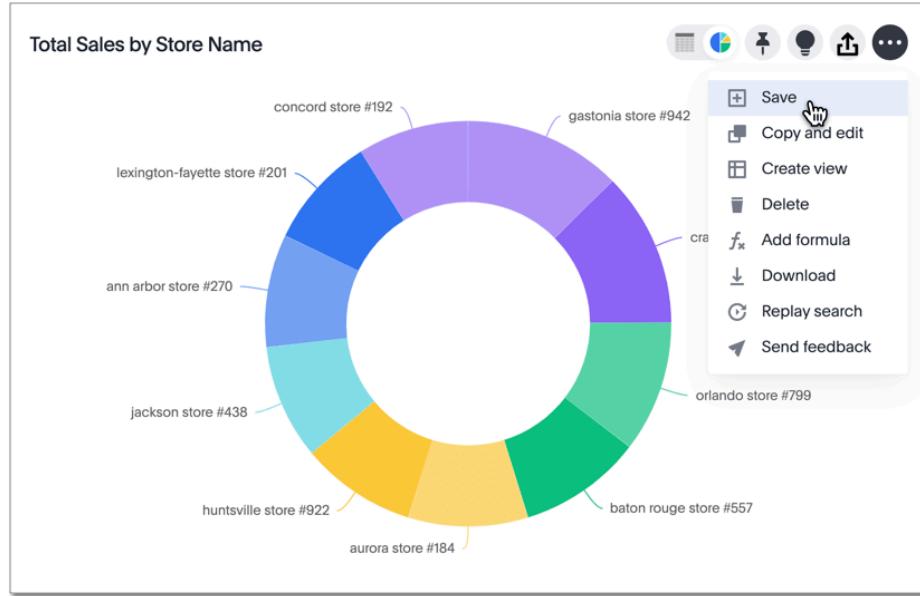
You do not have to be an administrator or the owner of an Answer to share saved Answers. Whenever you view an Answer, you have the option of sharing it with others. The Answer appears in its most recent state when you share it. For example, if you add a filter after saving the Answer and then share it, the Answer you share has that filter.

You can share an Answer from the list of Answers on the [main Answers page \[See page 336\]](#), or from the Answer itself [\[See page 338\]](#).

Share from the Answers page

To share an Answer from the **Answers** page, follow these steps.

1. Configure the Answer to look as it must appear when you share it.
2. Save the Answer by clicking the ellipsis icon , and selecting **Save**.



- Click **Answers** on the bar at the top of your screen.

Name	Stickers	Modified	Author
Copy of Sales by State - Last 3 Months		1 week ago	Retail
Product Name, Category Name, Reorder Level, Product Status, Supplier...		1 week ago	Inventory

- Select the Answer you want to share from the list of Answers by hovering over it and clicking the empty check box that appears.
- Click **Share**.

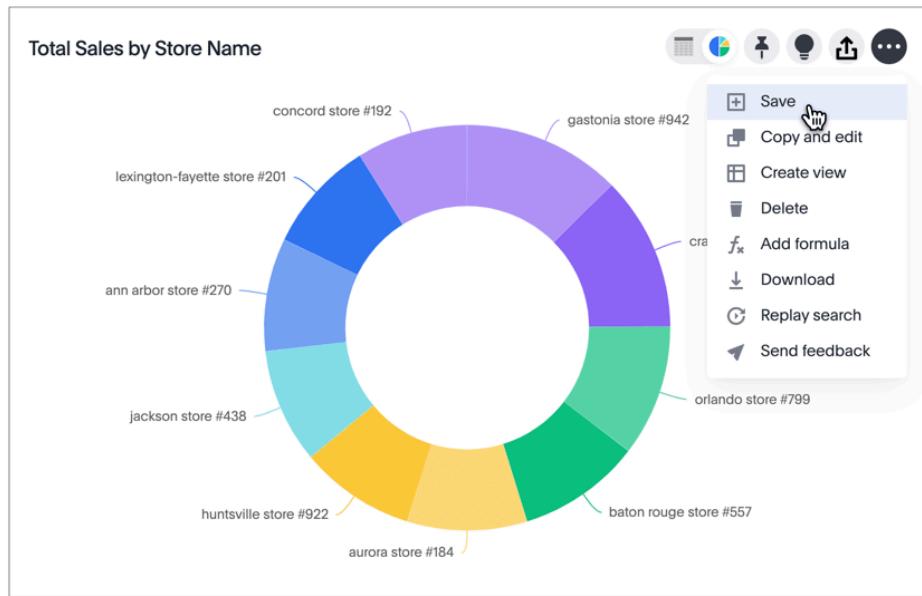
Name	Stickers	Modified	Author
<input checked="" type="checkbox"/> Copy of Sales by State - Last 3 Months		1 week ago	Retail
<input type="checkbox"/> Product Name, Category Name, Reorder Level, Product Status, Supplier...		1 week ago	Inventory

- Specify permissions. [See page 339]

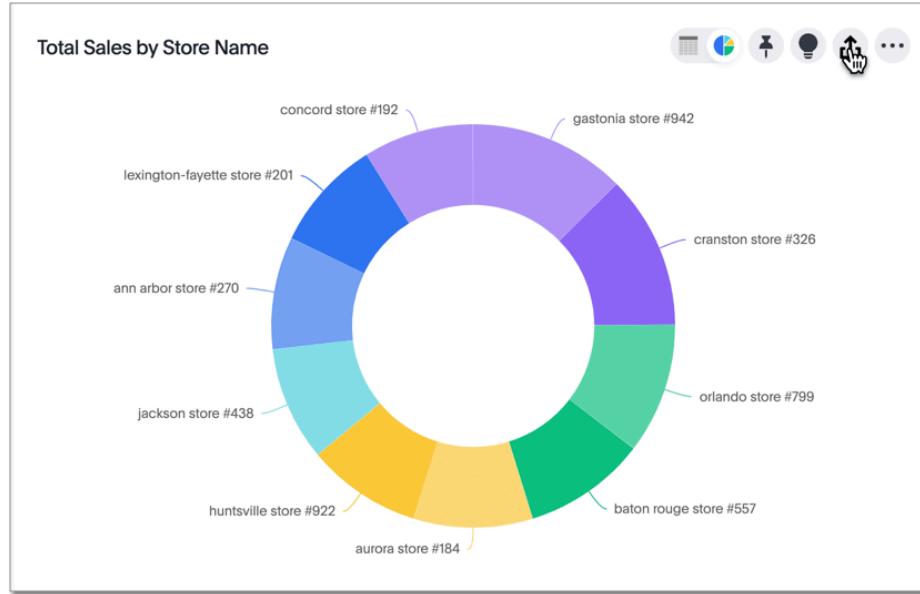
Share from within the Answer

To share an Answer from within the Answer, follow these steps.

1. Configure the answer to look exactly like it must appear when you share it.
2. Save the answer by clicking the ellipsis icon  , and selecting **Save**.



3. Share the answer by clicking the sharing icon .



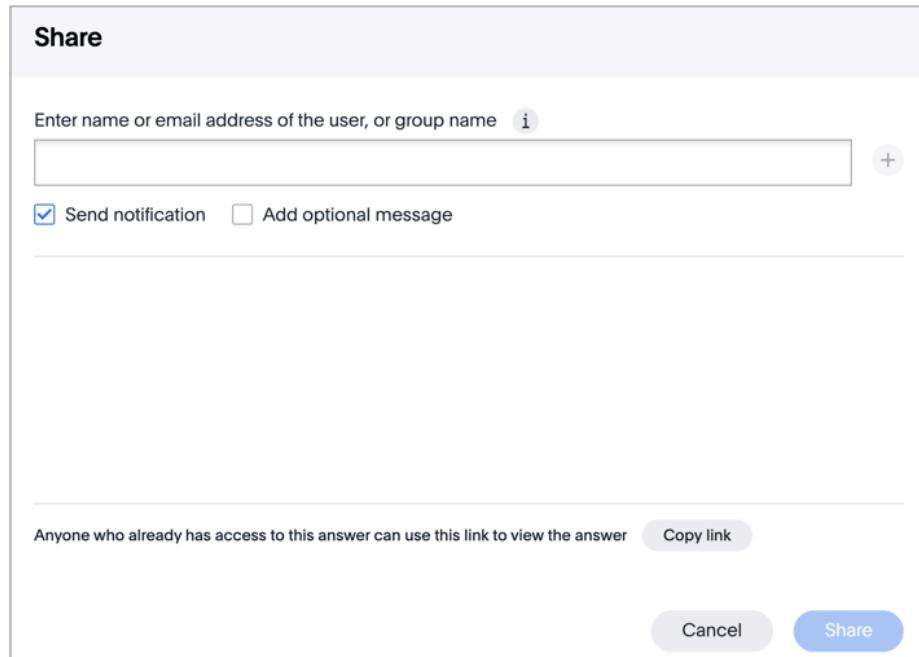
4. If you did not save the answer, ThoughtSpot prompts you to name and save the answer before sharing it.

The figure shows a modal dialog box with the title "Save the answer and share it". It contains two input fields: "Name the new answer:" and "Optional description". At the bottom right are two buttons: "Cancel" and "Save". A hand cursor is hovering over the "Save" button.

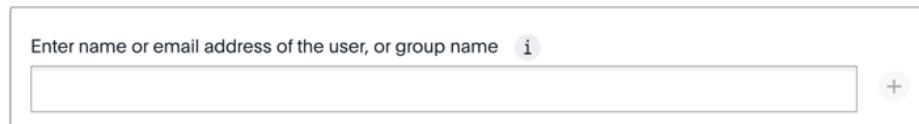
5. Specify permissions. [See page 339]

Specify permissions

1. After you click the **Share** icon, the sharing dialog box appears.



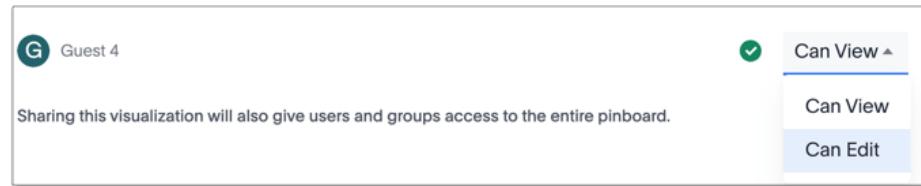
2. Enter users or groups with whom you want to share this object in the text box.



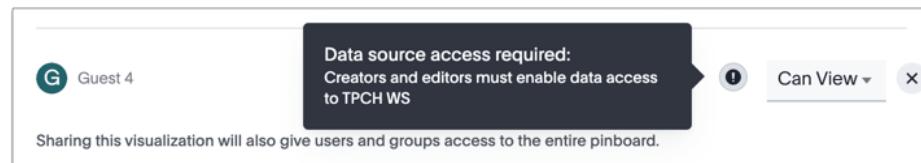
Note that you can only enter email addresses whose domains are in your list of allowed domains. These domains appear when you click on the info button **i**.

Tip: If you want to hide the allowed email domains for your company, or otherwise customize them, contact ThoughtSpot support.

3. Configure the level of access by selecting from the drop-down list. You can select:
 - **Can View** to provide read-only access. If the user doesn't have access to the underlying worksheet, they can only view the shared object.
 - **Can Edit** to allow modification. Enables renaming or deleting the shared object. If a user with edit privileges modifies a shared object, the object saves their changes.



4. If the selected group or user does not have access to the underlying data, you must enable access to the worksheet, view, or table. A black warning symbol appears when you try to share with a user who does not have underlying data access. If you click on it, it tells you to enable access:



If you own the underlying data source, you can enable access through the sharing dialog box. If you do not own the data source, ThoughtSpot emails the owner of the data source or your ThoughtSpot administrator to ask them to share the data.

To enable access, select **Give view access to underlying data sources** at the bottom of the dialog box. You can click on the arrow to view the data sources.

Share

Enter name or email address of the user, or group name [i](#)

Send notification Add optional message

 Guest 4  Can View ▾

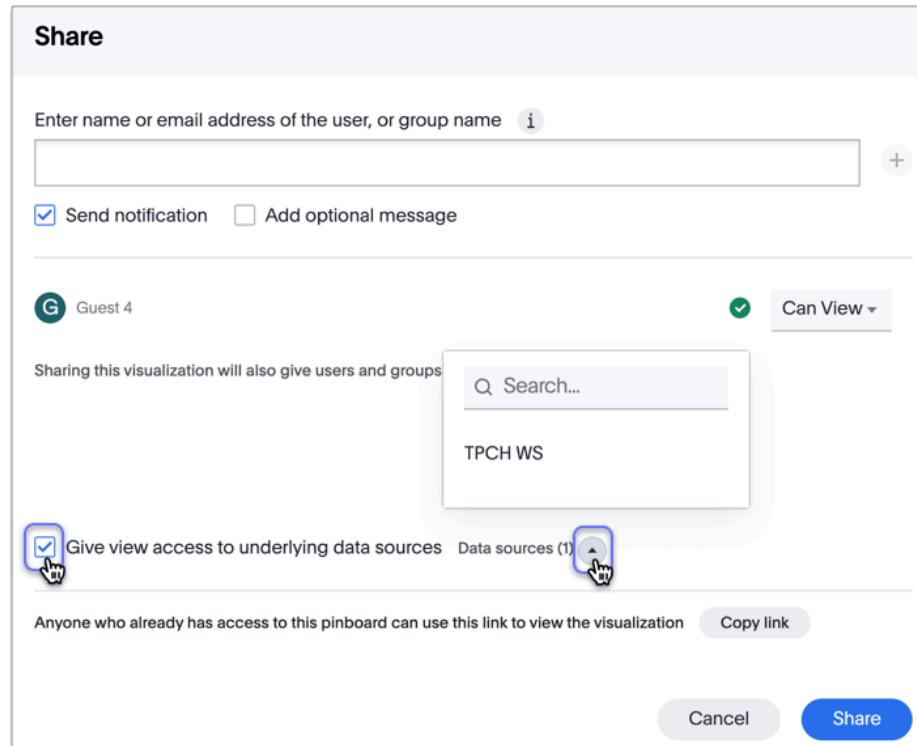
Sharing this visualization will also give users and groups:

Search...
TPCH WS

Give view access to underlying data sources [Data sources \(1\)](#) 

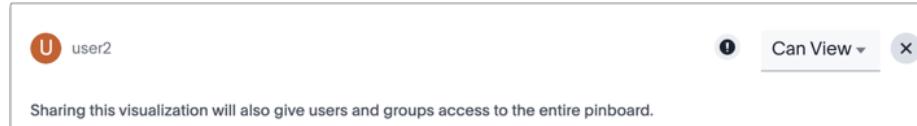
Anyone who already has access to this pinboard can use this link to view the visualization [Copy link](#)

[Cancel](#) [Share](#)

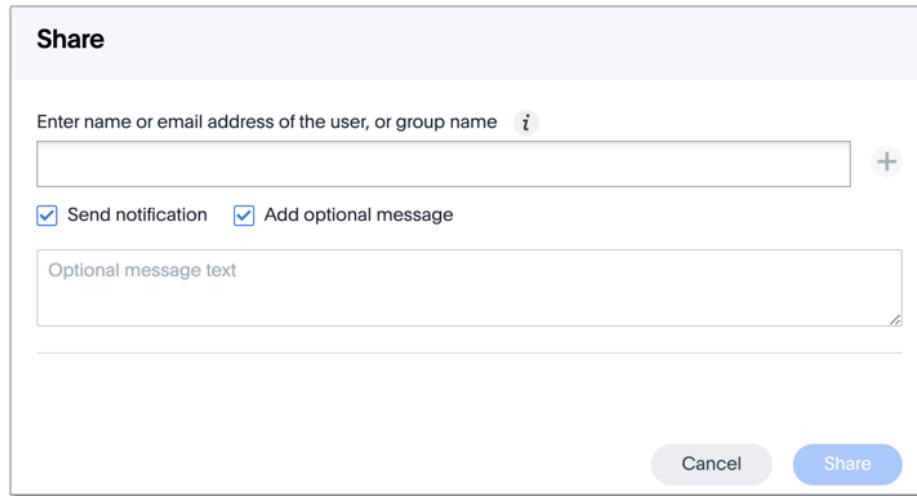


After you enable access, the warning symbol turns into a green checkmark.

5. To stop sharing with a user or group, click the x icon .

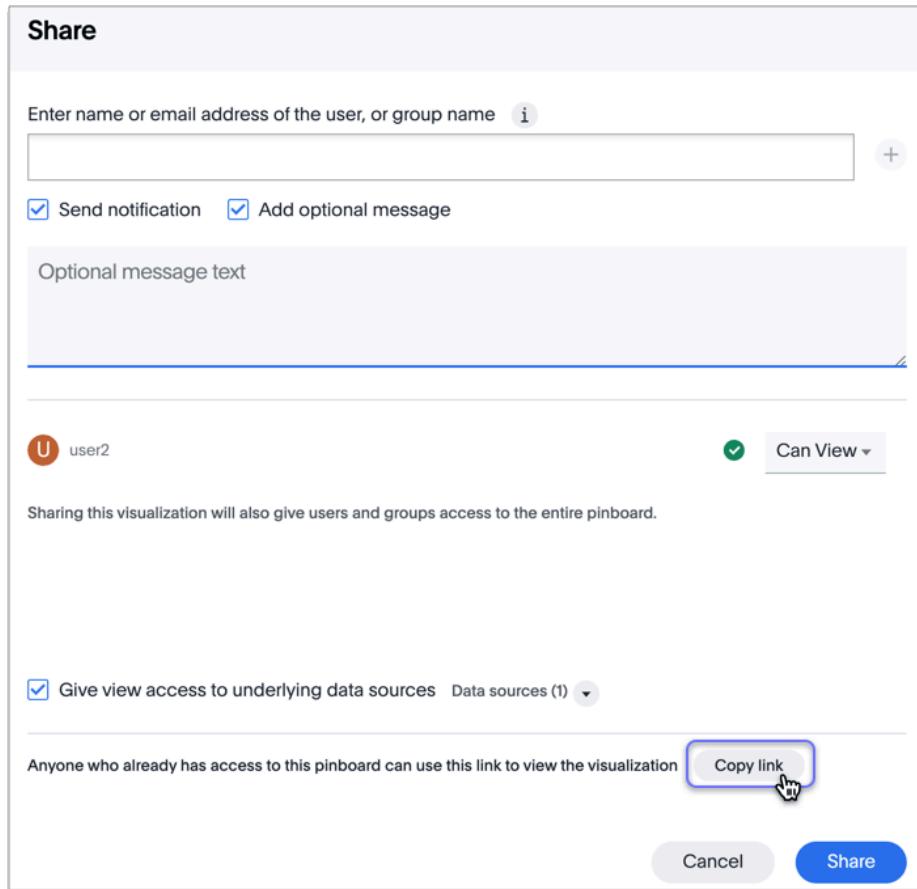


6. You can send an email notification and an optional message:



7. You can also copy a direct link to the Answer, Pinboard, or visualization within a Pinboard that you are sharing, and separately send that link to users after you share the object with them. Simply click the **Copy link** button at the bottom of the sharing dialog box.

Note that sending users this link does not share the object with them. You must also share the object by clicking the **Share** button at the bottom of the dialog box.



8. Click **Share**.

Share

Enter name or email address of the user, or group name [i](#)

Send notification Add optional message

Optional message text

 user2  Can View ▾

Sharing this visualization will also give users and groups access to the entire pinboard.

Give view access to underlying data sources [Data sources \(1\)](#) ▾

Anyone who already has access to this pinboard can use this link to view the visualization [Copy link](#)

[Cancel](#) [Share](#) 

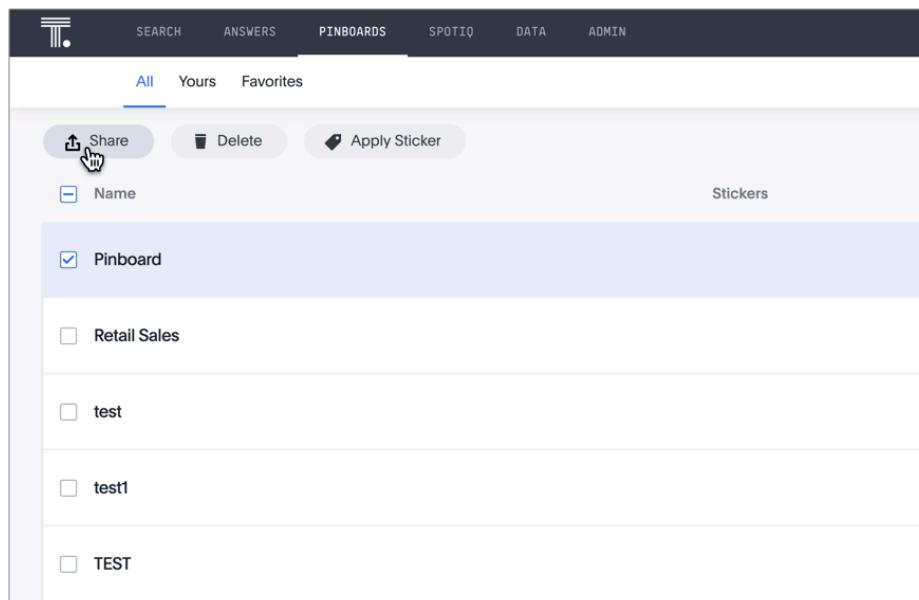
Revoke access (unshare)

Summary: Learn how to revoke access to an object.

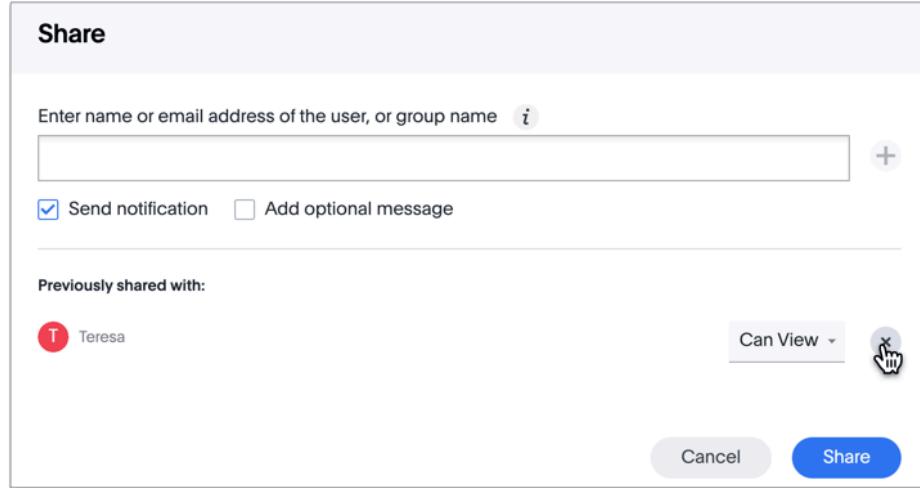
You may need to revoke access to an object (table, worksheet, pinboard, or answer) that you previously shared. Unsharing an object is very similar to sharing it.

Follow these steps to unshare one or more objects:

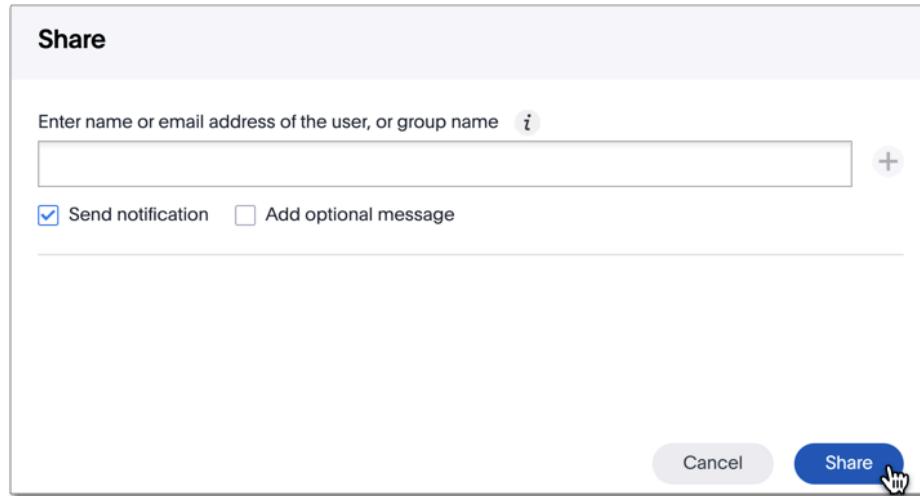
1. Navigate to the current section from the top menu bar.
 - If the object is a table or worksheet, click **Data**.
 - If the object is a pinboard, click **Pinboards**.
 - If the object is an answer, click **Answers**.
2. Hover over the object(s) you want to unshare, and select them by clicking the checkbox to the left of their name.
3. Click the **Share** icon.



4. Hover over a user or group with whom you want to stop sharing the object and click the **X** to remove them.



5. Click **Share** to update sharing permissions, and unshare with the specified users and groups.



6. The **Share settings updated** notification appears on the bottom of your screen.

Share settings updated. X

Security for SpotIQ functions

Summary: Learn about SpotIQ security.

SpotIQ is a feature in ThoughtSpot that automatically generates insights into system data. The feature works on all the data in your system, search queries, saved Answers, and Pinboards. Because SpotIQ uses the same data security model as other data in the system, there is no need to build a new schema or security model to support it.

Note: Insights are turned off by default. They do not appear on the Home page, or on Pinboards, unless your administrator turns on the feature.

SpotIQ automatically generates insights into data when a user requests them. These requests can be run immediately or users can schedule regular analysis. SpotIQ requests for insights can put additional load on your system depending on the amount of data being analyzed and how frequently users make use of it.

Users are required to have the **Has Spot IQ privilege** to use this feature. Users without this privilege cannot view insights unless the users that generated the insights add them to a pinboard and share the pinboard with others.

You may want to restrict access to a subset of your users or even to a subset of your managers. To restrict access to this feature:

1. Create a group called **SpotIQUsers**.
2. Grant this group **Has Spot IQ privilege**.
3. Enable users to this group.

About row level security (RLS)

Summary: Using row level security, you can restrict data that appears in search results and pinboards by group.

Row level security (RLS) allows you to restrict a group's access to table row data. You do this by creating a *rule* that associates a filter with a group. When a group member searches, views an answer, or otherwise works with data, ThoughtSpot evaluates the rules and prevents the display of the restricted data. Users see only the data they are permitted to see.

How does RLS impact user interactions?

The security rules apply to objects shared with users individually or through groups they are a member of. The rules restrict the visible data when users:

- view a table
- view a worksheet derived from the table
- search for data in the worksheet or table
- view answers from restricted data — either that they've created or that were shared with them
- interact with pinboards from restricted data — either that they've created or that were shared with them

Search suggestions also fall under row-level security. If a user would not have access to the row data, then values from the row do not appear in **Search** suggestions.

Why use RLS?

RLS allows you to set up flexible rules that are self-maintaining. An RLS configuration can handle thousands of groups. There are several reasons you might want to use row level security:

Reason	Example

Hide sensitive data from groups who should not see it.	In a report with customer details, hide potential customers (those who have not yet completed their purchase) from everyone except the sales group.
Filter tables to reduce their size, so that only the relevant data is visible.	Reduce the number of rows that appear in a very large table of baseball players, so that players who are no longer active are not shown except to historians.
Enable creation of a single pinboard or visualization, which can display different data depending on the group who is accessing it.	Create one sales pinboard that shows only the sales in the region of the person who views it. This effectively creates a personalized pinboard, depending on the viewer's region.

Related information

- To continue learning about RLS, see [How rule-based RLS works \[See page 351\]](#).
- **Search** suggestions relies on compile indices to present suggestions to users from your data. See [Manage suggestion indexing \[See page 428\]](#) to learn how to configure suggestions.

How rule-based RLS works

Summary: Use rule-based RLS to restrict a group's access to data.
Users see only accessible row data.

Row level security works at the group level and is configured on tables. A table's RLS rules also apply to any objects with data from that table. So, searches, answers, worksheets, and pinboards that rely on a table's data fall under RLS rules.

Worksheet queries and RLS

You cannot set RLS rules on worksheets, only on tables. However, administrators can disable RLS on worksheets that are derived from tables with RLS rules. After RLS rules are disabled, users with access to the worksheet can see all its data.

By default, worksheet queries only take into account RLS rules on tables whose columns appear in the query. Other related tables that may underly the worksheet are ignored. This means that not all RLS rules on underlying tables are applied when a user queries a worksheet.

You can configure a stricter application of RLS rules to take into account RLS rules from all the tables underlying the worksheet. This is recommended if you have key dimension tables that worksheets rely on but that are not necessarily regularly accessed through query. To do this, contact ThoughtSpot Customer Support.

Privileges that allow users to set, or be exempt from, RLS

Users in the **Administrators** group or with the **Has administration privilege** have full access to everything in the system. As a result:

- Row level security does not apply to them.
- They can create, edit, and delete RLS rules.
- They can also disable RLS rules on individual worksheets.

If your installation has enabled the **Can Administer and Bypass RLS** privilege, administrators can also grant **Can Administer and Bypass RLS** to groups. Members of groups with **Can Administer and Bypass RLS**:

- Are exempt from row-level security (RLS) rules.
- Can add/edit/delete existing RLS rules.
- Can check or uncheck Bypass RLS on a worksheet.

This behavior is true regardless of whether the privilege is from a direct group membership or indirect (through a group hierarchy).

Examples of RLS rules

An RLS rule evaluates against two system variables:

Function	Description	Examples
ts_groups	Returns a list of all the groups the current logged in user belongs to. For any row, if the expression evaluates to true for any of the groups, the user can see that row.	ts_groups = 'east'
ts_username	Returns the user with the matching name.	ts_username != 'mark'

ThoughtSpot filters a table's rows by evaluating a rule against the authenticated user.

A rule is an expression that returns a boolean, `TRUE` or `FALSE`. If the rule evaluates to `TRUE`, a user can see that row. If the rule evaluates to `FALSE` for the user, then the user cannot view the data and instead they see the message `No data to display`.

Rule expression can be implicit or explicit. And rules may or may not contain logic. A simple implicit RLS rule has the format:

`COLUMN_FILTER = ts_groups`

An example of an explicit rule that contains logic would be:

`if (COLUMN_FILTER) then true else false`

Rules can also reference tables other than the table you are securing.

Consider a simple RLS rule example. Your company has `vendor-purchase` table such as:

DATE	VENDOR	AMOUNT
12/11/39..	zendesk	116.00
12/11/39..	getquik com ca	289.70
12/11/39..	ikea	113.91
12/11/39..	costco	274.43
12/11/39..	waiters wheels pa	66.52
12/11/39..	waiters whee	76.49
12/11/39..	chipotle	175.33

You want to give your vendors the ability to see trends in company purchases. You give vendor personnel access to ThoughtSpot *and* add them to self-titled vendor groups. So, all users from the Starbucks vendor are in the `Starbucks` group and all users from `round table` are in the `Round Table` group. Then, you set a **Row security** on the `vendor-purchase` table as follows:

```
VENDOR = ts_groups
```

Only users in `Starbucks` group see `starbucks` data and so forth. Rules ignore case inconsistencies and spaces are evaluated so `round table` in the data matches the `Round table` group but not a group named `RoundTable`.

Rules can be simple or they can incorporate logic such as `if/then` rules. For example, vendors should see their own data but your accounts payable group needs to see all the vendor data:

```
VENDOR = ts_groups or 'Accounts Payable' = ts_groups
```

This rule continues to work as you add data from new vendor or team members to `Accounts Payable`. In this way, a well-written rule is *self maintaining*, meaning you don't have to revisit the rule as your system changes.

You can also create rules that reference tables other than the table you are securing. For example, if you have a `sales` table and `store` dimension table, you can use attributes from the `store` table to secure the `sales` table.

Multiple rules and multiple group membership

You can define multiple rules on table. In this case, ThoughtSpot treats the rules as additive. That is, they are applied using an `OR` operator. If any of the rules evaluate to `true` for a user on a row, that row's data is visible.

If a user is a member of multiple groups, the user can see all the rows that are visible to all of their groups. The most permissive policy is used.

Members of groups with **Can Administer and Bypass RLS** are exempt from row-level security (RLS) rules. This is true regardless of whether the group membership is direct or indirect (through a group hierarchy).

Best practices for using Rule-Based Row Level Security

Use these best practices for Rule-Based Row Level Security:

- Use **Share** as the first level of data access.

Non-administrative users and groups have no way to access any data without first having it shared with them. So, only share what you need.

When you share, share worksheets. This is a general best practice. Worksheets simplify the data environment for end users; they only need to choose among a few sources, rather than many tables. Also, one worksheet can also combine data from several tables.

- Set row level security wherever you want to keep data secure.

It is always a possibility that a particular search only includes data from a single table, and a user will see something they shouldn't. So, protect your data by setting row level security wherever you want to keep data secure.

- Explicitly grant access for users that should see all rows.

As soon as you define a rule on a table for one group, you prevent access by all others outside of that group hierarchy. Subsequent rules should specifically add groups that need access.

- Keep in mind that multiple rules on a table are additive with `or`.

If you are concerned with security, start with very limited access. Then, expand the access as needed.

- Keep rules simple.

Complex rules can impact the system performance. So, err on the side of simple rules rather than complex rules with a lot of logic.

Related information

- To learn the procedure you follow for setting a rule, [Set RLS rules \[See page 356\]](#)
- For a list of operators and functions you can use to build RLS rules see [Row level security rules reference \[See page 0\]](#).
- For information on bypassing rules on a worksheet, see [Change inclusion, join, or RLS for a worksheet \[See page 481\]](#).

Set row level security rules

Summary: Learn how to set RLS rules.

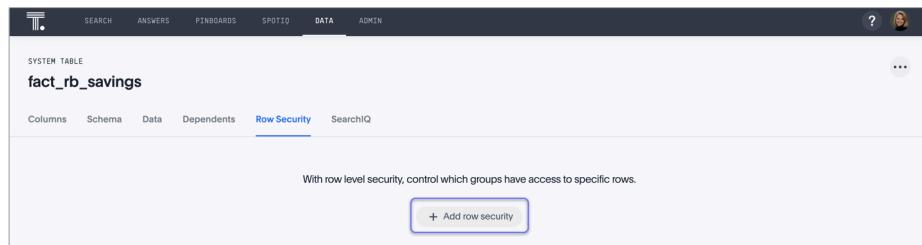
When rule-based row level security (RLS) is set, it prevents users from seeing data they shouldn't in tables and the objects derived from them. You must have administrative rights on ThoughtSpot to set RLS rules.

Before you create a rule, make sure you have read [How rule-based RLS works \[See page 351\]](#).

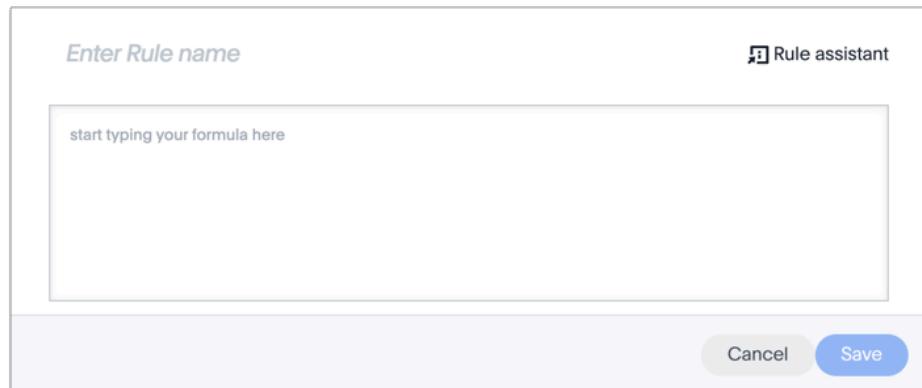
Create a rule on a table

You can set RLS rules *only* on tables. To set up rule-based row level security, do the following:

1. Click **Data**, and double-click a table.
2. Click **Row security**.
3. Click **+ Add row security**.



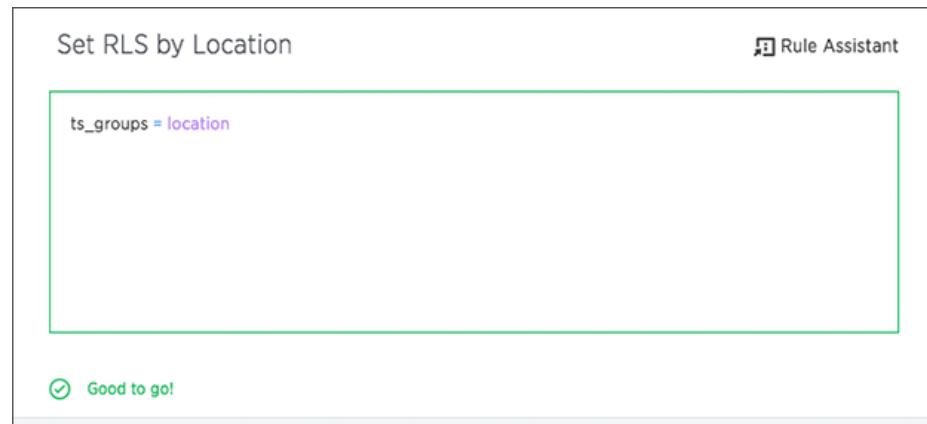
The system displays the Rule Builder.



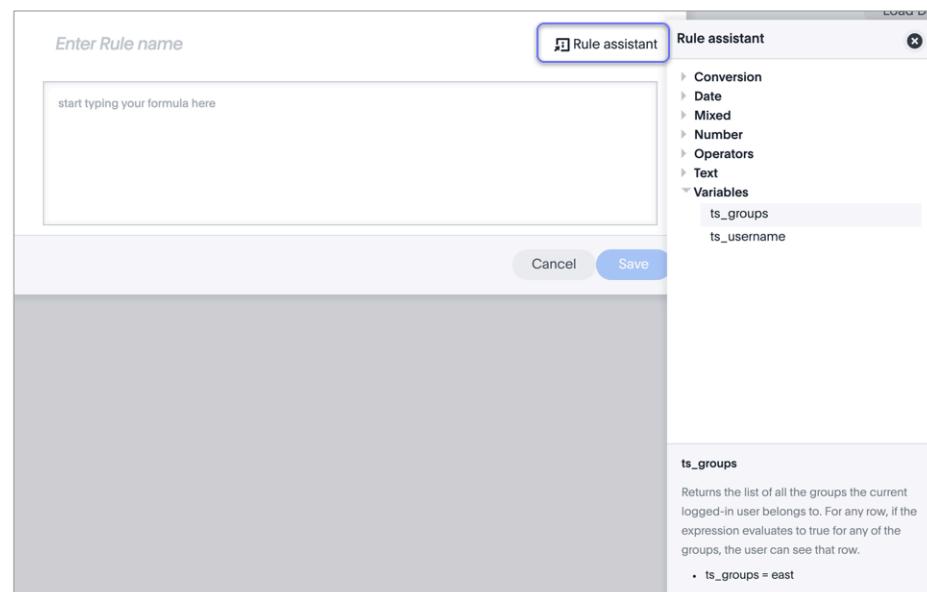
You define row level security by creating an expression that gets evaluated for every row and group combination. This powerful feature can be used with up to thousands of groups.

4. Open the Rule Builder.
5. Give your rule a name.
6. Enter an expression for your rule.

The rule gets evaluated against an authenticated user for every row and group combination. If the rule evaluates to `FALSE`, the user cannot see that row's data. Instead, they see the message `No data to display.` Use the variable `ts_groups` to refer to the group name.



You can see a list of available operators by clicking on **Rule Assistant**.



As you type, ThoughtSpot suggests formula syntax, variables, and column names. If you can't remember the exact column name or variable you want to use, the suggestions can help.

When your expression is valid, a green indicator appears at the bottom of the Rule Builder.

7. Click **Save**.

The rule you created is listed in the rules. You can edit the rule or add more rules by clicking **+ Add**.

Test your rule with restricted and unrestricted users

To test your rule, log in as users in different groups. Search within the table for data that your test user can and can't access. Make sure your test users can see the appropriate rows.

Related information

- Administrators can bypass the RLS rules set on a the table at the worksheet level. See [change the join rule or RLS setting for a worksheet \[See page 481\]](#).
- For a list of operators and functions you can use to build RLS rules, see the [row level security rules reference \[See page 0\]](#).

ThoughtSpot Lifecycle

Summary: This topic covers security processes for the entire lifecycle of a ThoughtSpot deployment from development, release, installation, upgrades, to software patching.

Overview

A ThoughtSpot deployment consists of the following high level software systems:

- Operating System (OS) and software packages installed on the OS
- Third-party software
- ThoughtSpot application services (binaries and configuration)

Operating System

All ThoughtSpot physical appliances, virtual machines (VMs) and public cloud images come pre-installed with CentOS 7. The [CentOS](https://www.centos.org/) (<https://www.centos.org/>) distribution of Linux is owned by [RedHat](https://www.redhat.com/en) (<https://www.redhat.com/en>) and closely tracks versions of RedHat Enterprise Linux (RHEL).

ThoughtSpot uses the minimal install of CentOS 7 with the addition of a few software packages (e.g. Python) needed for ThoughtSpot operations. The most notable change to the installation is to the Linux kernel, which is sourced from the current long term stable kernel version instead of the default included in CentOS 7 (kernel-lt package). To list all the installed packages, see [Checking Package Versions \[See page 362\]](#).

Third-Party Software (Middleware)

Third party software used includes Java, Boost C++ libraries, Google protocol buffers, etc. These are software components necessary for operation of the ThoughtSpot application. ThoughtSpot only uses software licensed for distribution.

Development and Release Process

ThoughtSpot releases its software as a tarball containing all the ThoughtSpot application (binaries and configuration), third-party software, and an operating system image. Installation or update using this release tarball on appliances, VMs, or cloud instances updates each of these components.

Operating System

Building the operating system image including software packages is a multi-step process:

1. Begin with the set of packages in the base OS image and our added packages.
2. Configure all installation to only use official public RedHat repositories.
3. For each package, install the current stable version including any security patches.
4. Bring up the image on all supported platforms for stability and performance testing along with the ThoughtSpot application stack. Success criteria: no OS impact on stability or performance.
5. Scan the Operating System and ThoughtSpot application stack using Qualys scans with additional modules enabled: Vulnerability Management, Web App Scanning.
6. Review all vulnerabilities found. Success criteria is zero severity 4+ vulnerabilities.
7. Assuming all above testing and exit criteria are met, the OS image is considered qualified.

Third-Party Software

Third-party software is periodically sourced from the upstream distribution of each software component. Unlike OS and ThoughtSpot application, this changes less frequently and on an as needed basis, when any new security vulnerability or stability issue is discovered in the library. The list of all third-party software as well as licensing details are here.

ThoughtSpot Application

ThoughtSpot follows industry standard best practices for writing robust software. Every code change is reviewed by at least one engineer. Our engineering team consists of senior engineers from Enterprise software and web companies.

ThoughtSpot uses a small number of proven programming languages powering some of the largest enterprises in the world. ThoughtSpot tracks stability, performance, and reliability of our software and services aggressively. The ThoughtSpot platform is trusted by dozens of global F2000 organizations.

Protection of Source Code

Source code is private and not shared publicly, e.g. all distribution to customers is in binary or minified format to discourage reverse engineering.

Automated Tools

We use automated tools and infrastructure like Jenkins, Kubernetes, AWS, partnering with the teams behind these systems so as to adopt best practices. For example, all our automation runs through Jenkins, which is managed by CloudBees (the company behind Jenkins) using an enterprise license with regular security patching, and so on. We upgrade our automation tools regularly.

Independent Testing

Independent testing is done outside of the product team by pre sales and post sales before promoting to production. Some areas are tested by third party testers.

Security Hardening

Starting 4.5.1.5, we have also taken specific steps to incorporate most of CIS standard recommendations towards hardening.

Installation and Upgrade Process

ThoughtSpot is installed or updated from a release tarball which contains the ThoughtSpot application (binaries and configuration), third-party software, and Operating System image.

Operating System Image Installation

Installing ThoughtSpot on any node (VM, cloud instance, appliance) automatically updates the operating system and required packages on the node. No Internet or repository access is required for this, the update is applied directly from the release tarball.

Specifically, all nodes running ThoughtSpot are required to have two root partitions on their boot drive of which one of them is booted from at any given time. During installation or update, the Operating System image contained in the release tarball is copied into the second currently-unused root partition and the system switches to it through a reboot.

Checking OS Package Versions

The following command run from any ThoughtSpot node will indicate versions of all installed packages:

```
rpm -qa
```

Upgrades

ThoughtSpot patches the Operating System at the time of upgrades. The exact same process used during installation is also applied during upgrades. The previous OS image on a node gets replaced by the new image carried in the release tarball.

Only some releases may patch the Operating System, not all. Typically, all major and minor releases (e.g. 4.4, 4.5, 4.5.1, 5.0) upgrade OS patches, whereas only some patch releases (e.g. 4.4.1.4) contain OS patches.

Distributed Clusters and Failure Handling

On distributed clusters, individual nodes receive the OS image from the release tarball individually.

Initially, the new image is deployed on a single node only. When that node is deemed healthy following the update and a rich set of tests, the image is made available to remaining nodes in the cluster.

If a node fails to patch, then ThoughtSpot support will modify the upgrade workflow to either retry the patching or skip and exclude the node.

Third-Party Software

Installation or upgrade of ThoughtSpot deployments automatically upgrades all third-party software to the version included in the release tarball.

Security Scanning and Patching Process

The ThoughtSpot Security team continuously scans security bulletins for new vulnerabilities discovered in included OS packages (e.g., Linux Kernel, libc) and third party software (e.g., Java). Additionally, weekly scans are done for all release branches using Qualys with the following additional modules enabled: Vulnerability Management, Web App Scanning. The security scans discover vulnerabilities at all layers: OS, third-party software, as well as ThoughtSpot application binaries and configuration. Additionally, ThoughtSpot periodically scans all source code for third-party software as well as ThoughtSpot's proprietary code base for vulnerabilities or unsafe usage using SourceClear.

After a critical new vulnerability is found (severity 4 or 5), ThoughtSpot includes the corresponding patch in the next patch release for all supported release branches. Consult ThoughtSpot documentation or support to find out if you are on an active or supported release branch.

After a new patch release with a critical security vulnerability is available, customers are encouraged to upgrade their deployment quickly.

Latency

We recommend customers to wait for the next regular release for receiving security patches. However, should a critical vulnerability be discovered in the interim, ThoughtSpot can push out a new patch release containing the required patches, if available upstream.

ThoughtSpot targets a three week or less cadence for generating patch releases for all supported release branches. Timeline for the new release and patching depends on availability of the patch upstream (e.g., not all vulnerabilities in Linux are immediately fixed) and qualification (ThoughtSpot qualifies each build on each supported cloud and on-prem platform). If a fix is unavailable upstream at the moment, customers and ThoughtSpot support can work together to identify potential workarounds.

Storage Security

Encryption at Rest

- On-prem: Not supported yet
- Cloud: Supported on AWS (See <https://aws.amazon.com/>), GCP (See <https://cloud.google.com/%2F%5D>), Azure (See <https://azure.microsoft.com/en-us/>)

Secure Erase

Current erase guide (See https://thoughtspot.egnyte.com/dl/E1eYDyfotL/SOP-520-0007-00-User-Data-Removal.pdf_)

Encryption of data in transit within a cluster

Summary: ThoughtSpot can encrypt data in transit within a cluster.

Overview

ThoughtSpot supports encryption of data in transit within a cluster (traffic flowing between multiple nodes in a cluster). Encryption in transit within the cluster is primarily needed for cloud deployments of ThoughtSpot. This is accomplished using IPSec.

IPSec operates in two modes: tunnel mode and transport mode. ThoughtSpot recommends using **transport mode** to set up IPSec for encrypting in-transit data.

Note: While IPSec provides additional security, it also reduces network bandwidth between nodes.

ThoughtSpot supports IPSec encryption using strongSwan (an open-source IPSec-based VPN solution for Linux and other UNIX based operating systems).

Summary:

1. Use IPSec in Transport mode for host-to-host IPSec communication.
2. Use the strongSwan package.
3. Use AES-GCM for ESP protocol (since it provides authenticated encryption and provides better network bandwidth).

Note: IPSec is supported in ThoughtSpot software versions starting from 4.5.1.4

Deployment

1. **Enabling IPSec:** Run following command on any ThoughtSpot node after cluster has been configured and right built has been deployed: `# tscli ipsec enable`.
2. **Disabling IPSec:** Run following command on any node of the cluster: `# tscli ipsec disable`.
3. **Checking status of IPSec:** Run following command on any node of the cluster `# tscli ipsec status`.

4. IPSec configuration and settings are NOT persistent across cluster backup/restore and will have to be re-enabled.
5. **Adding a node in an IPSec enabled cluster:** IPSec settings are automatically configured across nodes as new nodes are added to a ThoughtSpot cluster that has IPSec enabled.
6. **Removing a node in IPSec enabled cluster:** IPSec settings are not impacted when nodes get removed from a ThoughtSpot cluster that has IPSec enabled. If any failure occurs during node removal, IPSec would need to be re-enabled by using the following command: `tscli ipsec enable`.
7. **Manually enable/disable IPSec on a single node(if any IPSec related failure occurs on that node):**

Manually enable and start strongSwan:

```
sudo systemctl enable strongswan  
sudo systemctl start strongswan
```

Manually stop and disable strongSwan:

```
sudo systemctl stop strongswan  
sudo systemctl disable strongswan
```

Note: You can check `/etc/strongswan/ipsec.conf` and `/etc/strongswan/ipsec.secrets` on each node to verify that they are consistent across all nodes.

Firewall configuration

The following ports must be open between nodes to allow IPSec encryption:

- UDP port 500 (for IKE)
- UDP port 4500 (for IPSec over IDP)
- IP Protocol 50 (ESP)

System administration

Summary: Learn tools to help administer ThoughtSpot.

System administration includes applying upgrades, backing up and restoring the cluster, snapshotting, and adding or removing nodes.

Administration tools

Use these tools to perform administrative actions:

- [tscli \[See page 0\]](#): an administrative command line interface.
- [tsload \[See page 0\]](#): a command for loading data directly into the database.
- [TQL \[See page 0\]](#): a command line SQL interface to interact with databases.

Send logs when reporting problems

Summary: You can report problems to ThoughtSpot support or your administrator by sending logs.

You can generate a log bundle which you can then send to ThoughtSpot Support or you can send logs direct to your administrator.

Generate log bundle

Before you can send a log bundle to ThoughtSpot Support, you must [Connect to the ThoughtSpot Support file server \[See page 139\]](#). This is a one-time setup operation.

To generate a log bundle:

1. Log in to the Linux shell using SSH.
2. Issue the command to generate the log bundle:

```
tscli callhome generate-bundle  
--d <directory> --since <num_of_daysd>
```

Note: Don't forget to include `d` after your specified number of days. For example, `30d`.

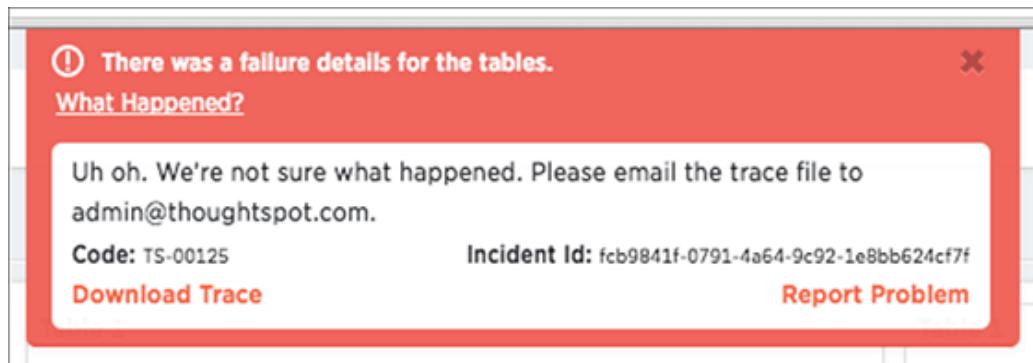
3. Change directories to the directory where you wrote the log bundle.
4. Issue the command to send the log bundle to ThoughtSpot Support:

```
tscli fileserver upload  
--file_name <file>  
--server_dir_path <path>
```

Send a log to the administrator

Alternately, you can easily send log files directly to your administrator with a single click. When ThoughtSpot encounters a problem, a red bar displays in the browser with an error message. You can use the **Report Problem** option to complete this task.

Click **Report Problem** in the bottom right corner of the error message.



The logs will be sent to your administrator as an email attachment from your email account. Your administrator then has the option to followup with ThoughtSpot, if necessary.

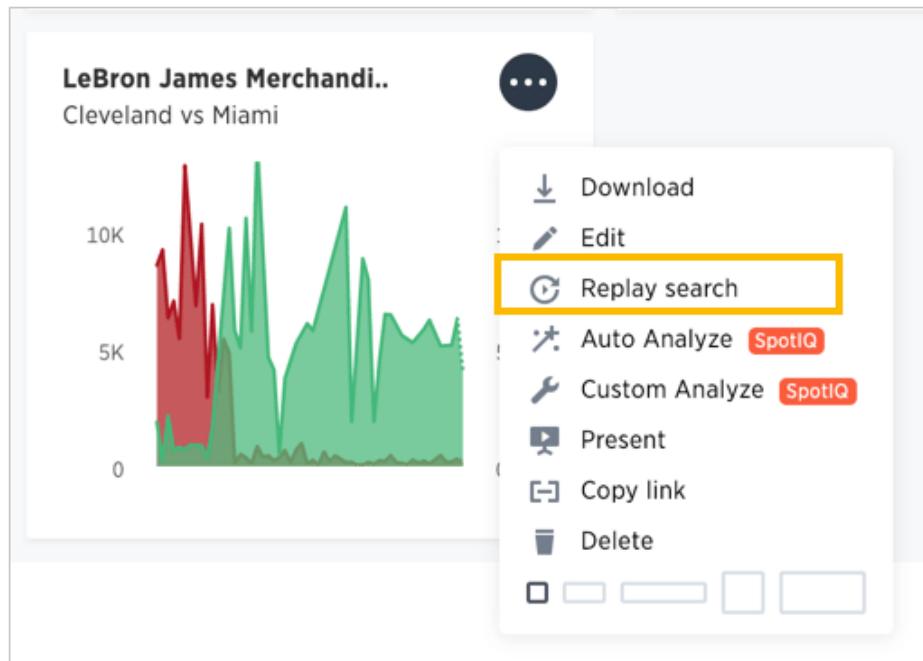
Set up recording for Replay Search

Summary: You can record a search replay to create training for your users on how to search your own data.

Recording a search replay requires administrator privileges and a Firefox browser. You must override some of your browser security settings in order to use the ThoughtSpot application to make the recording. This is a one time setup operation. If you do not wish to do this, you can replay the search and record it using QuickTime, Camtasia, or another screen cam recording tool.

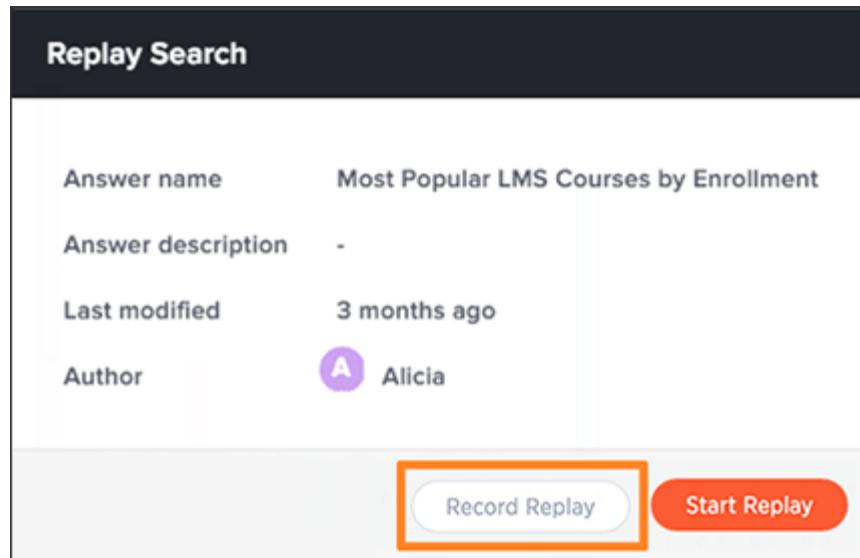
To record a search replay using ThoughtSpot:

1. While viewing a chart or table in ThoughtSpot, click the **Replay Search** icon.



2. Click the **Record Replay** button.

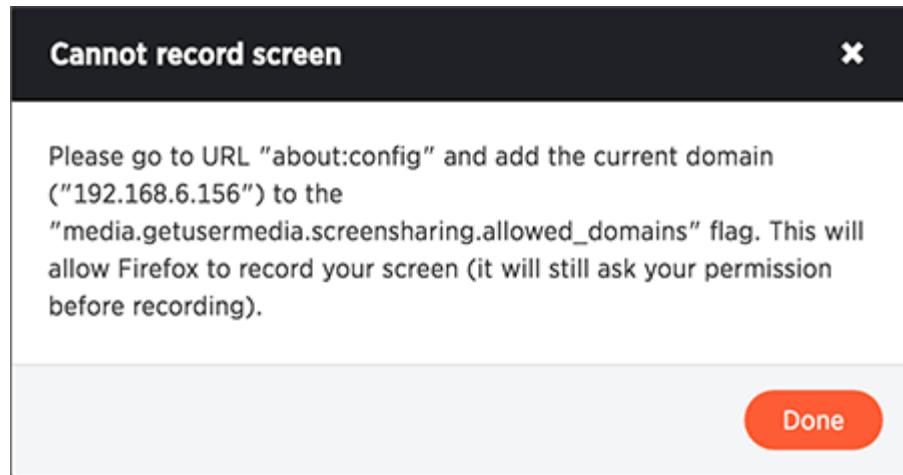
If you do not see the button, you must log in as a user with administrator privileges.



A message will display, showing a URL and a domain or an IP address.

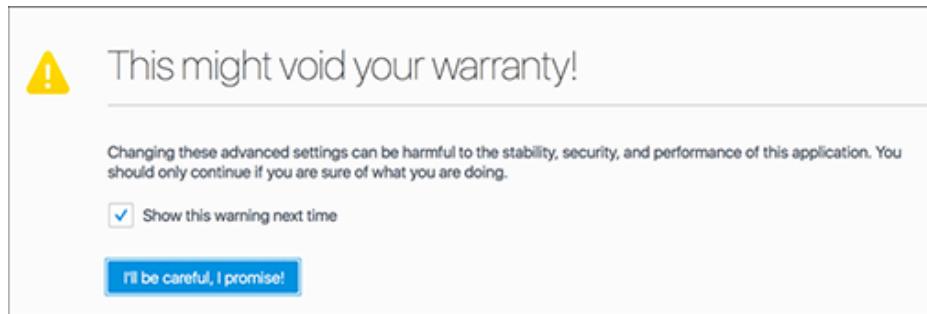
3. Make note of both of these items.
4. Open a new browser tab and go to the URL shown in the message (for example, "about:config").

Depending on which browser and version you are using, you may need to access the browser configurations through a menu or by typing in a different URL. Check your own browser help section for information on how to access the browser configuration settings, if necessary.



You may see a message warning that you are about to override the browser settings.

5. If you trust yourself, click “I'll be careful, I promise!”.

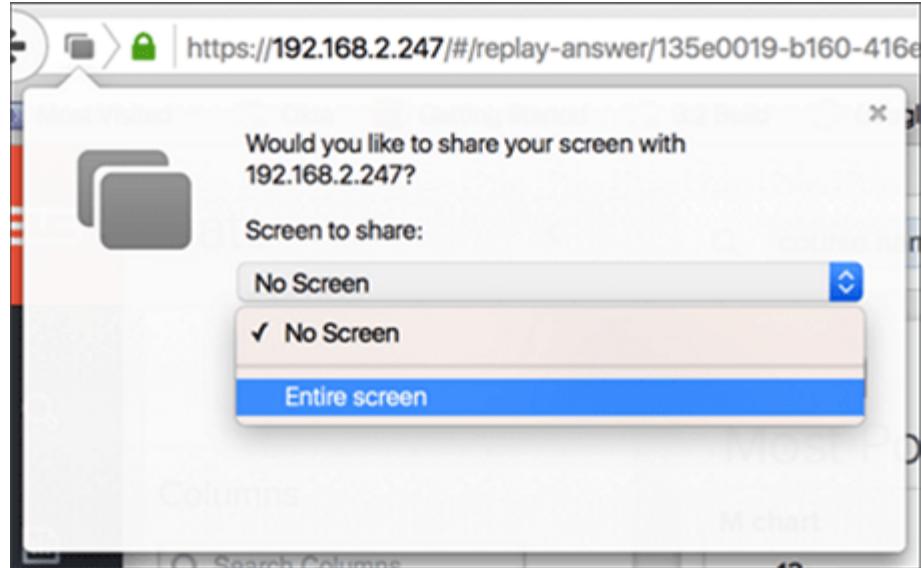


6. Find the setting for **media.getusermedia.screensharing.allowed_domains**, and add the domain used by ThoughtSpot.

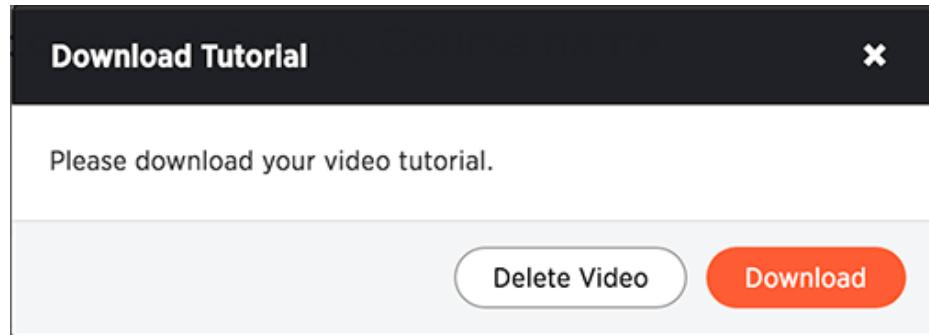
This domain will be the same one you made note of from the **Cannot record screen** message.

media.getusermedia.agc_enabled	default	boolean	false
media.getusermedia.audiocapture.enabled	default	boolean	false
media.getusermedia.browser.enabled	default	boolean	true
media.getusermedia.noise	default	integer	1
media.getusermedia.noise_enabled	default	boolean	true
media.getusermedia.playout_delay	default	integer	10
media.getusermedia.screensharing.allow_on_old_platforms	default	boolean	false
media.getusermedia.screensharing.allowed_domains	default	string	webex.com,*.webex.com,ciscospark.com,*
media.getusermedia.screensharing.enabled	default	boolean	true
media.gmp-gmpopenh264.abi	user set	string	x86_64-gcc3-u-i386-x86_64
media.gmp-gmpopenh264.lastUpdate	user set	integer	1454453226

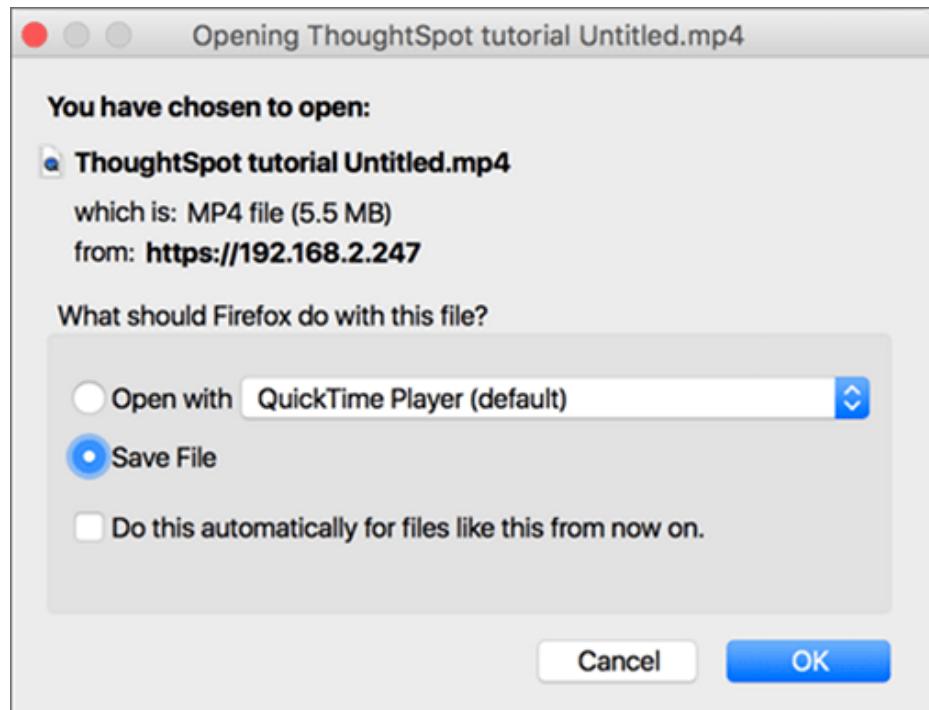
7. If you see a message asking if you'd like to share your screen with the IP address or domain name of ThoughtSpot, select **Entire screen**.



8. When the search replay has been recorded, you can see a confirmation. Select **Download**.



9. Save the recording on your computer by selecting **Save File** and clicking **OK**.



Upgrade a cluster

Summary: Contact ThoughtSpot Support to upgrade a cluster to a new release.

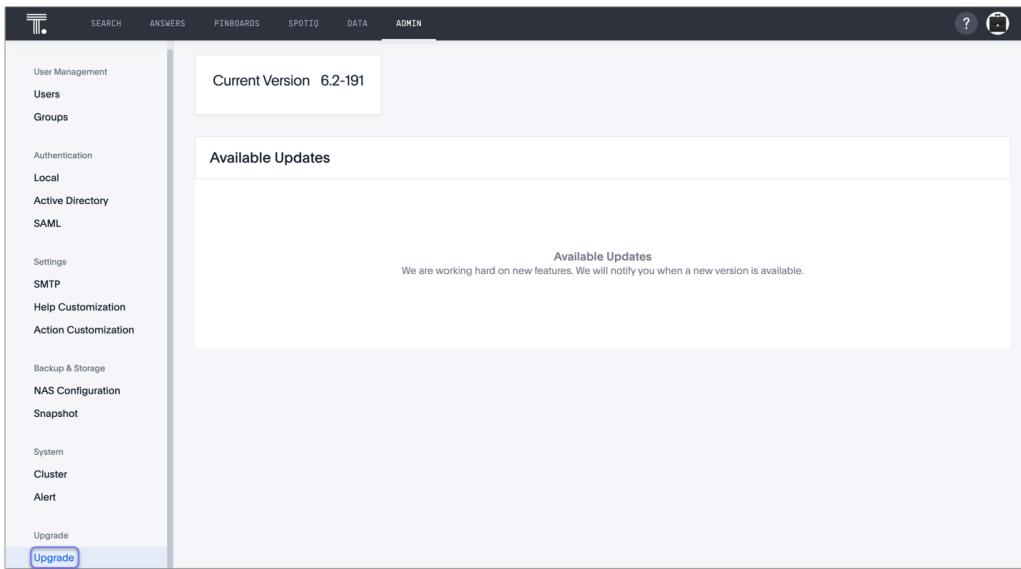
ThoughtSpot is installed or updated from a release tarball that contains the ThoughtSpot application (binaries and configuration), third-party software, and Operating System image. Third party softwares are licensed software components necessary for operation of the ThoughtSpot application. These include Java, Boost C++ libraries, Google protocol buffers, and so on.

ThoughtSpot patches the Operating System at the time of upgrades. Installation and upgrade use the same process, replacing the older OS image on a node by the new image delivered in the release tarball.

ThoughtSpot Support will contact you to schedule an update when a minor or major upgrade becomes available.

You can also view available releases from the Admin Console. To view new releases, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Upgrade** from the side navigation bar that appears.

The **Upgrade** page shows your cluster's current version, and any new releases that are available. If no releases are available, you see the following screen.



If releases are available, and you would like to upgrade, [contact ThoughtSpot Support \[See page 0\]](#) to upgrade your cluster.

Understand the backup strategies

Summary: Consider the strategies for backing up your ThoughtSpot cluster.

Snapshots

A snapshot is a point-in-time image of your running cluster. Snapshots are both taken on and restored to a cluster while it is running. Each cluster has a periodic snapshot configuration enabled by default. This configuration instructs the system to periodically take snapshots. Creation of a snapshot could take as little as 20 seconds, but depends on the number of objects in your cluster. After creation, a snapshot persists on disk in the cluster's HDFS.

You can also create a snapshot manually. You should create a snapshot before making any changes to your cluster's environment, loading a large amount of new data, or changing the structure of a table. A snapshot may only be restored to the same cluster on which it was taken. The cluster software release version must match the snapshot release version.

If you need to move data between clusters or restore to a cluster that was updated to a new release, contact [ThoughtSpot Support](#) [See page 0].

Backups

A backup is a procedure that stores a snapshot *outside* of a ThoughtSpot cluster. Backups are stored in a directory on a local or network file system. You can store all of the data associated with a snapshot, a portion of that data, or only metadata.

There is no default configuration enabled for backing up a cluster. You can configure a periodic backup policy yourself, or you can take backups manually. Backing up periodically protects your company from losing data and/or user work.

You can use a backup to restore a cluster to a prior state or to a differently configured appliance. You can also use a backup to move a cluster from an appliance to a virtual cluster, or vice versa.

Offline backup cluster

The most robust strategy for backup and recovery requires having a backup cluster offline that is kept in sync with the production cluster. Then, if the production cluster fails, the backup cluster can be drafted to take its place with minimal loss of work and disruption to operations.

Details on this architecture, and instructions on setting it up, are available in the ThoughtSpot Disaster Recovery Guide, which you can request from ThoughtSpot.

Choosing a strategy

Depending on your situation and your goals, you can choose to use either a snapshot or a backup. This table should help you decide:

	Snapshot	Backup
Purpose	Restore a cluster to particular point in time.	<ul style="list-style-type: none">• Restore a cluster to a prior state.• Move a cluster to a different hardware, cloud, or VMware appliance.• Remove a node.• Restore to a cluster that runs a different release from the one where the backup was taken.
Storage	In the cluster's HDFS	Options: <ul style="list-style-type: none">• Outside the cluster on a local disk• Outside the cluster on an NAS disk• You can back up an AWS cluster using an S3 bucket.• You can back up a GCP cluster using a GCS bucket.

Advantages	<ul style="list-style-type: none">• Can be taken on, or restored to, a running cluster• Fastest create and restore	<ul style="list-style-type: none">• Very stable.• Can be used to recover from data loss or corruption, even if the cluster is destroyed.• Can be <i>full</i>, <i>lightweight</i>, or <i>dataless</i>.
Limitations	<ul style="list-style-type: none">• Includes all data, state, and metadata created between snapshot create and restore.• Snapshots do not copy over anything that is in the home directories or root partitions of an instance. If you routinely add flat files or scripts directly, make separate copies of these flat files and scripts.• Lost if the HDFS name node fails, if you lose multiple disks, or if the entire cluster is destroyed• Can be restored only to the cluster on which the snapshot was taken	<ul style="list-style-type: none">• Requires deleting the existing cluster first.• You are responsible for validating your backup configuration as viable for restoring a cluster.• Backups do not copy over anything that is in the home directories or root partitions of an instance. If you routinely add flat files or scripts directly, make separate copies of these flat files and scripts.• Best practice recommends you to maintain multiple backups.• Typically, very large in memory size.

You should never restore from a snapshot or backup yourself. Contact [ThoughtSpot Support \[See page 0\]](#) for help.

Additional resources

As you develop your expertise in backups and snapshots, we recommend the following ThoughtSpot U course:

- [Snapshots and Backups \(\)](https://training.thoughtspot.com/3-snapshots-backups/461810)

See other training resources at



Understand backup/snapshot schedules

Summary: Learn about backup and snapshot schedules.

You can schedule periodic snapshots and backups. ThoughtSpot comes configured with a default periodic snapshot policy. For backups, there is no such policy. You may want to configure one or several of your own backup policies. This section helps you understand existing schedules and how to configure new schedules.

Configuration format

ThoughtSpot uses a [protocol buffer](https://developers.google.com/protocol-buffers/) (<https://developers.google.com/protocol-buffers/>) configuration file to hold snapshot and backup policies. There are slight differences between the configuration of snapshots and backups. Refer to [work with snapshots \[See page 385\]](#) and [configure periodic backups \[See page 399\]](#). However, the file format defines a `schedule` structure, which is the same for both snapshots and backups. The following example shows the `schedule` format:

```
schedule {  
    period {  
        number: integer  
        unit: MINUTE | HOUR | DAY  
    }  
    retention_policy {  
        bucket {  
            time {  
                number: integer  
                unit: MINUTE | HOUR | DAY  
            }  
            capacity: integer  
        }  
    }  
    offset_minutes_from_sunday_midnight: integer  
}
```

The `schedule` has the following components:

period	Specifies the frequency in the chosen <code>unit</code> . You can specify the <code>unit</code> as <code>MINUTE</code> , <code>HOUR</code> , or <code>DAY</code> .
retention_policy	Specifies retention intervals. Retention is on a first-in-first-out (FIFO) basis. So, the system discards the oldest result. You can specify the <code>unit</code> as <code>MINUTE</code> , <code>HOUR</code> , or <code>DAY</code> . You can specify multiple retention buckets and they can have different retention policies.
offset_minutes_from_sunday_midnight	Determines the minute within the hour you'd like execution to start. Setting this to zero is equivalent to midnight.

Work through an example schedule

In this section, you work through an example schedule. This working example is the actual default snapshot schedule set on every ThoughtSpot instance.

```
schedule {
    period {
        number: 1
        unit: HOUR
    }
    retention_policy {
        bucket {
            time {
                number: 1
                unit: HOUR
            }
            capacity: 3
        }
        bucket {
            time {
                number: 4
                unit: HOUR
            }
            capacity: 2
        }
    }
    offset_minutes_from_sunday_midnight: 0
}
```

Under this policy, the system takes a snapshot every hour, starting at midnight. You can see that by combining the `period` of 1 hour with the midnight offset of 0.

```
schedule {  
    period {  
        number: 1  
        unit: HOUR  
    }  
    ...  
}  
offset_minutes_from_sunday_midnight: 0  
}
```

Using this frequency, a total of 24 snapshots are taken in a day.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	...	24
---	---	---	---	---	---	---	---	---	----	----	----	----	----	-----	----

If you were to specify a `number` of 2, the frequency changes. The system creates the first snapshot at midnight, and creates subsequent snapshots every 2 hours:

2	4	6	8	10	12	14	...	24
---	---	---	---	----	----	----	-----	----

You use the `retention_policy` to control how many snapshots are kept. In this example, the first bucket retains a snapshot every three hours.

```
retention_policy {  
    bucket {  
        time {  
            number: 1  
            unit: HOUR  
        }  
        capacity: 3  
    }  
    ...  
}
```

At the beginning of the fourth hour, the system discards the snapshot from the first hour. This is in accordance with FIFO behavior. Therefore, this retention bucket contains snapshots from hours 2, 3, and 4.

1D	2R	3R	4R	5	6	7	8	9	10	11	12	13	14	...	24
----	----	----	----	---	---	---	---	---	----	----	----	----	----	-----	----

The second bucket retains the snapshot taken at four hour intervals.

```
retention_policy {  
    ...  
    bucket {  
        time {  
            number: 4  
            unit: HOUR  
        }  
        capacity: 2  
    }  
}
```

It retains two of these four-hour-interval snapshots at any one time. By hour 9 during the day, you have the snapshots from hour 4 and hour 8 in this second bucket.

1	2	3	4R	5	6	7	8R	9	10	11	12	13	14	...	24
---	---	---	----	---	---	---	----	---	----	----	----	----	----	-----	----

What is in the first bucket in hour 9? The first bucket, with number 1 and capacity 3, has the snapshots from hour 9, 8, and 7.

At the end of the day, in the first bucket, you have the 22nd, 23rd, and 24th snapshot. In the second bucket, you will have the 20th hour and the 24th hour snapshots.

1	...	12	13	14	15	16	17	18	19	20R	21	22R	23R	24R
---	-----	----	----	----	----	----	----	----	----	-----	----	-----	-----	-----

What if you changed the period frequency to every 2 hours? What is in your buckets at hour 24?

1	...	12	14	16	18R	20R	22R	24R
---	-----	----	----	----	-----	-----	-----	-----

When defining a policy, it can be helpful to graphically represent the frequency you configure. Then, determine which time blocks are important to retain before determining your retention bucket.

Work with snapshots

Summary: A snapshot is a point-in-time image of your running cluster. You can use a snapshot to restore the cluster to a specific point in time. In this section, we describe how to work with the default snapshot configuration that is enabled on every cluster, and how to make manual snapshots.

To work with snapshots, use the `tscli` command line interface.

⚠ Warning: Backups rely on the snapshot system, so you must never disable the periodic snapshot system. If you disable periodic snapshots and enable periodic backups, the backups either fail or use an outdated snapshot.

About manual snapshots

You must create a snapshot [through the Admin Console \[See page 386\]](#) or [using tscli \[See page 388\]](#) before making any changes to your cluster's environment, loading a large amount of data, or changing the structure of a table. ThoughtSpot supports up to 20 manual snapshots. You must clear them to create new snapshots. You can delete a snapshot [from the Admin Console \[See page 387\]](#).

ThoughtSpot generates a snapshot in approximately 20 seconds, depending on the size of the cluster. To restore from a snapshot, contact [ThoughtSpot Support \[See page 0\]](#).

⚠ Note: During an upgrade, all snapshots from the previous version of ThoughtSpot become manual snapshots.

Manage snapshots through the Admin Console

From the Admin Console, you can [create manual snapshots \[See page 386\]](#), [delete existing snapshots \[See page 387\]](#), and view the details for all your existing snapshots.

To manage or create snapshots using the Admin Console, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Snapshot** from the side navigation bar that appears.

Snapshot Name	Scheduled On	Snapshot Style	Status
_x_periodic_20200720T140000_706402	20/07/2020 07:00	PERIODIC	Completed
_x_periodic_20200721T200000_386470	21/07/2020 13:00	PERIODIC	Completed
_x_periodic_20200722T100000_138090	22/07/2020 04:00	PERIODIC	Completed
_x_periodic_20200722T150000_138753	22/07/2020 08:00	PERIODIC	Completed
_x_periodic_20200722T190000_139608	22/07/2020 12:00	PERIODIC	Completed
_x_periodic_20200722T230000_140311	22/07/2020 16:00	PERIODIC	Completed
_x_periodic_20200723T030000_141146	22/07/2020 20:00	PERIODIC	Completed
_x_periodic_20200723T050000_141538	22/07/2020 22:00	PERIODIC	Completed
_x_periodic_20200723T190000_462055	23/07/2020 12:00	PERIODIC	Completed

You can view snapshot details, such as **name**, **style** (manual or periodic), **status**, and the **date and time** a snapshot was scheduled.

Create manual snapshots in the Admin Console

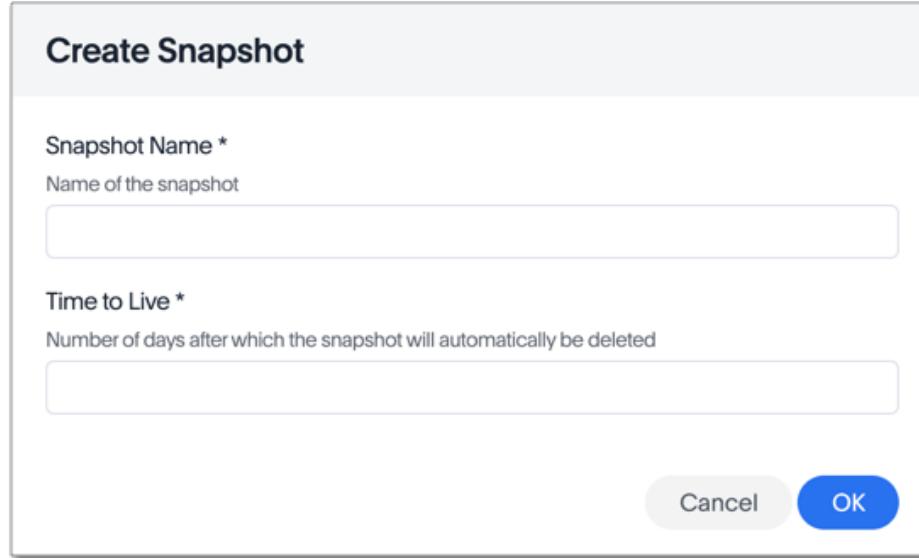
To create a manual snapshot in the Admin Console, follow these steps.

1. Click **Create** at the upper left side of your screen.

Snapshot Name	Scheduled On	Snapshot Style	Status
_x_periodic_20200720T140000_706402	20/07/2020 07:00	PERIODIC	Completed
_x_periodic_20200721T200000_386470	21/07/2020 13:00	PERIODIC	Completed
_x_periodic_20200722T100000_138090	22/07/2020 04:00	PERIODIC	Completed
_x_periodic_20200722T150000_138753	22/07/2020 08:00	PERIODIC	Completed
_x_periodic_20200722T190000_139608	22/07/2020 12:00	PERIODIC	Completed
_x_periodic_20200722T230000_140311	22/07/2020 16:00	PERIODIC	Completed
_x_periodic_20200723T030000_141146	22/07/2020 20:00	PERIODIC	Completed
_x_periodic_20200723T050000_141538	22/07/2020 22:00	PERIODIC	Completed
_x_periodic_20200723T190000_462055	23/07/2020 12:00	PERIODIC	Completed

2. Specify a **name** and a **time to live** for the snapshot.

If you do not want the system to delete the snapshot automatically after a specified number of days, enter `-1` as the **time to live**.

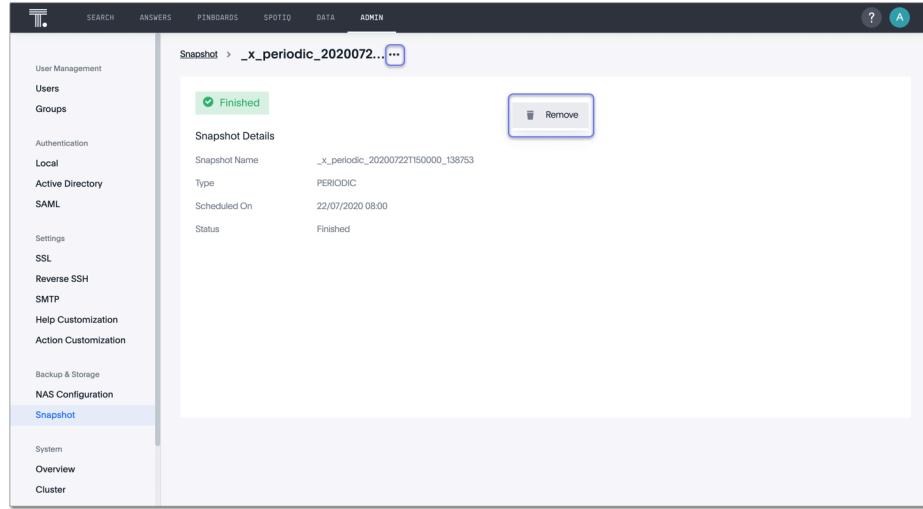


3. Click **OK**.

Delete snapshots in the Admin Console

To delete a snapshot in the Admin Console, follow these steps.

1. Click on the snapshot you would like to delete.
2. Select the **more options** menu icon  next to the snapshot name.
3. Click **remove**.



Create a snapshot using tscli

To create a snapshot using `tscli`, follow these steps:

1. Log in to the Linux shell using SSH.
2. Create a snapshot with the command `tscli snapshot create`, specifying the following parameters:
 - `name` : Specify a name for the snapshot. Note that snapshot names must not exceed 44 characters.
 - `reason` : Specify a reason for creating the snapshot.
 - `ttl` : Specify the snapshot's time to live, in days. If you specify `ttl` as `4`, the system deletes the snapshot after four days.

```
$ tscli snapshot create <name> <reason> <ttl>
```

3. Check that the snapshot was created successfully by checking the directory listing:

```
$ tscli snapshot ls
```

Your snapshot should appear in the list of snapshots that `tscli snapshot ls` generates.

Configure periodic snapshots with snapshot policy

Each ThoughtSpot cluster automatically creates periodic snapshots, based on its default snapshot policy. The default snapshot policy is enabled for every cluster. You can display the current policy for periodic snapshots. You can also change this policy.

View snapshot policy

To view your current periodic snapshot policy, follow these steps:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Enter `tscli snapshot-policy show <name>` command to view the policy.

In the following example, the policy starts at midnight on Sunday. It retains the snapshots from the previous three 4-hour intervals, and two snapshots from two of the previous 4-hour intervals. Here, ThoughtSpot retains 7 periodic snapshots overall. For detailed information about understanding the schedule, see [Understand backup/snapshot schedules \[See page 380\]](#).

```
$ tscli snapshot-policy show
schedule {
    period {
        number: 1
        unit: HOUR
    }
    retention_policy {
        bucket {
            time {
                number: 1
                unit: HOUR
            }
            capacity: 3
        }
        bucket {
            time {
                number: 4
                unit: HOUR
            }
            capacity: 4
        }
        bucket {
            time {
                number: 1
                unit: DAY
            }
            capacity: 4
        }
        bucket {
            time {
                number: 1
                unit: WEEK
            }
            capacity: 2
        }
    }
    offset_minutes_from_sunday_midnight: 0
}
enabled: true
```

Update the snapshot policy

Do not change the default policy, unless it is recommended by [ThoughtSpot Support \[See page 0\]](#).

Note: Your policy must hold no more than 20 snapshots at any time. Exceeding this number impacts cluster performance.

To update your current periodic snapshot policy, follow these steps:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Enter `tscli snapshot-policy update`.

```
$ tscli snapshot-policy update
```

3. In the editor, make changes to the current policy, and save them.

You cannot delete the snapshot policy. However, you can either disable or enable it.

Disable the snapshot policy

To disable your current periodic snapshot policy, follow these steps:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Enter `tscli snapshot-policy disable`.

```
$ tscli snapshot-policy disable
```

Enable the snapshot policy

To enable a specific periodic snapshot policy, follow these steps:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Enter `tscli snapshot-policy enable`.

```
$ tscli snapshot-policy disable
```

Understand backup modes

Summary: Learn about types of backups.

A backup is a procedure that stores a snapshot *outside* of a ThoughtSpot cluster. Backups are stored in a directory on a local or network file system. You can store all of the data associated with a snapshot, a portion of that data, or only metadata. Other advanced administrative operations also use backups.

You can use a backup to restore a cluster to a prior state or to a differently configured appliance. You can also use a backup to move a cluster from an appliance to a virtual cluster, or vice versa.

You can create a manual backup or configure an automated, periodic backup. For manual backups, the system creates a backup using the named snapshot you specify. For periodic backups, the system uses the most recent snapshot to create the backup.

Warning: You should never disable the periodic snapshot system, because backups rely on it.

For example, if you have disabled the periodic snapshots system and periodic backups are enabled, then the periodic backup may use a very outdated snapshot or it may fail all together.

ThoughtSpot usually stores backups on a [NAS \(network attached storage\) file system](#) [See page 132] but you can store them on a local disk as well. You can back up an AWS cluster [using an S3 bucket](#) [See page 0]. You can back up a GCP cluster [using a GCS bucket](#) [See page 0]. When creating a backup, ThoughtSpot copies a release tarball and several supporting files to a disk you specify. Storing these supporting files takes about 10 GB of extra space beyond the backup itself. The final backup image is smaller because these extra files are removed after the backup completes successfully. So, make sure you have enough disk space both to *take* a backup and store the result. Use the `tscli storage df` command to identify the amount of space available.

You can create a backup using one of three modes: `full`, `lightweight`, or `dataless`.

Full backups

Full backups are entire backups of the cluster with all data, whether loaded from the web interface or from `tsload`. This is the best mode for restoring a cluster and all your data. After a `FULL` backup is created, you can move the backup between clusters, even if the cluster configuration is different. Full backups can be as large as 20 GB in addition to the 5 GB of additional files. Some installations can exceed these limits, this is why it is important to test your backup configuration.

Before creating a manual backup or configuring automated backups, make sure there is enough disk space on the target disk. Consider an example, where you want to store three backups. If the backup itself takes 18GB, you need about $18 + 5 = 23$ GB of free disk space. Don't forget that the backup size can grow over time, so you should occasionally check to ensure you are not in danger of running out of disk space to store backups.

Lightweight backups

Lightweight backups contain everything that makes up a cluster:

- Cluster configuration (SSH, LDAP, etc.)
- In-memory data cache
- All data that is stored unencrypted in HDFS
- Data uploaded by users
- Metadata for the data store
- Users, groups and permissions
- Objects created by users (pinboards, worksheets, and formulas) with their shares and permissions.
- Data model and row-level security rules.

Lightweight backups do **not** contain data loaded through ThoughtSpot Loader (`tsload`), ODBC/JDBC drivers, and Data Connect. The expectation is that data loaded by `tsload` is from external sources and so can be re-loaded after the cluster is restored. An exception is if these mechanisms were used to load data into tables that were first created through CSV import (that is, a user first loaded the tables using the UI). In this case, the data, like the tables they were loaded into, are saved.

Dataless backups

A dataless backup saves a backup of the schema (metadata), with no data. Dataless backups allow you to send a copy of your cluster metadata to ThoughtSpot Support for troubleshooting without compromising data security and privacy.

When restoring from a dataless backup, you must supply the correct release tarball, since this type of backup does not include the software release.

Create a manual backup

Summary: Learn how to manually create a backup.

ThoughtSpot enables you to manually create a backup to restore a cluster to a prior state or a differently configured appliance. You can also move the backup from an appliance to a virtual cluster or vice versa.

To restore a cluster from a backup, [contact ThoughtSpot Support \[See page 0\]](#).

You create a manual backup from an existing snapshot. So, you must identify an existing snapshot to use or take a new snapshot first. The time required to take a backup depends on the data size. Taking a backup does not take long, and happens in the background while the cluster is running.

Beginning with release 6.1, you can back up to an Amazon S3 bucket.

Using tscli

To manually create a backup using ThoughtSpot's command line interface, tscli:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. [Create a manual snapshot \[See page 388\]](#) or find a snapshot you want to use. To find a snapshot you want to use for your backup, run `tscli snapshot ls`:

```
$ tscli snapshot ls
```

```
Name          : pre330
Reason        : pre3.3.0
Hdfs snapshot: pre330
Start         : Wed May 4 18:07:32 2016
End           : Wed May 4 18:08:23 2016
Size(Full)    : 13.24 GB
Size(LW)      : 4.96 GB
Size(Dataless): 39.76 MB
```

```
...
```

3. Make sure you have enough room on the target disk.

In addition to the size of the snapshot, you must have 10 to 12 GB of disk space. This is because the process requires space for temporary files. You can use the `df -h` command to check disk size.

```
$ df -h
```

4. Create the backup, designating the [type of backup \[See page 393\]](#), the snapshot name, and a directory:

Choose the [mode of backup \[See page 393\]](#) you want to create, either `full`, `lightweight`, or `dataless`. You specify an existing directory path, but the folder (the last part of the path: `/home/admin/<folder>`) **must not** already exist. ThoughtSpot creates the folder when it runs a backup. The BASE value is the name of the backup.

```
$ tscli backup create [-h]
  [--mode {full|light|dataless}]
  [--type {full}]
  [--base snapshot_name]
  [--storage_type {local|nas|cloud}] [--remote]
  [--bucket_name]
  <name> <directory>
```

Here is a sample `tscli backup create` command:

```
$ tscli backup create --mode dataless <my_snapshot> <directory_path>
```

Backing up a ThoughtSpot cloud deployment

For details on how to back up a ThoughtSpot cluster hosted in the cloud, see:

- [Backup and Restore with AWS \[See page 0\]](#)
- [Backup and Restore with GCP \[See page 0\]](#)

Configure periodic backups

Summary: Learn how to configure automatic periodic backups.

You can configure ThoughtSpot to backup automatically at specified times. The policy allows you to control the type, frequency, retention periods (first-in-first-out), and output location for a periodic backup.

A periodic backup uses the same steps as creating a backup manually. However, you do not have to specify a snapshot name. The system uses the most recent snapshot. You can backup to a local file system, [mount a NAS \(network attached storage\) file system \[See page 132\]](#) to hold the backup, or [back up to an S3 bucket \[See page 0\]](#). A NAS file system is recommended. Make sure you have adequate space to store the number of backups you want to archive.

Default policy format

This is the format for a backup policy. Note that the command populates several parameters with their defaults. The default `mode` is `FULL`, the `type` is `STANDALONE`, and the `storage_type` is `NAS`. You can change the `mode` and `storage_type`, and you must specify a `directory` and a `name`.

```
name: "name_for_backup"
param {
    mode: FULL | DATALESS | LIGHTWEIGHT
    type: STANDALONE
}
schedule {
    period {
        number: integer
        unit: MINUTE | HOUR | DAY
    }
    retention_policy {
        time {
            number: integer
            unit: MINUTE | HOUR | DAY
        }
        capacity: integer
    }
}
offset_minutes_from_sunday_midnight: integer
}
directory: "NAME"
storage_type: NAS | LOCAL | S3
enabled: true (you only need this parameter if your storage_type is S3)
bucket_name: "your-S3-bucket-name-in-quotes" (You only need this parameter if your storage_type is S3. The bucket name must be in quotes.)
```

❶ Note: To periodically back up your cluster to an S3 bucket, specify S3 for the storage_type, and include the enabled and bucket_name parameters.

Before you begin

Before creating a policy, make sure you have read [Understand backup/snapshot schedules \[See page 380\]](#) for information on configuring a `schedule` element. In addition, you must specify the following parameters:

Element	Description
name	Specify a name for your backup, so you can identify it later.
mode	The backup mode. FULL backups are necessary for restoring a cluster. See Understand backup modes [See page 393] for details on each backup mode.
type	Only STANDALONE is supported.
directory	The location on the disk to place the backup. You specify an existing directory path, but the folder (the last part of the path: /home/admin/<folder>) must not already exist. ThoughtSpot creates the folder when it runs a backup.
storage_type	The type of storage you are using. NAS storage is recommended for FULL backups.

Create a backup policy

Backups cannot start when another backup is still running. Choose a reasonable frequency for the backup mode in your policy. For example, a FULL backup takes longer than a DATALESS backup. Consider the load on the system when configuring. Do not back up when the system would experience a heavy load. For example, you may want to take FULL backups late in the evening or on weekends.

The retention system deletes the oldest stored backup and the corresponding snapshot on a first-in-first-out basis (FIFO). This means that if you set a bucket retention of 1, the system stores a single backup at any one time. The system deletes the older backup after the new full backup is successful.

Configure using tscli

To configure periodic backups using the tscli:

1. Log in to the Linux shell using SSH.

```
$ ssh admin@<cluster-IP>
```

2. Find a directory with enough disk space to support the retention_policy number you configure.

You can use df -h to see free disk space and tscli snapshot ls to view existing snapshots and their size on disk.

3. Use the `tscli backup-policy create` command.

```
$ tscli backup-policy create
```

The command opens a `vi` editor for you to configure the backup policy.

4. Write and save the file to store your configuration.

By default, newly created policies are automatically enabled. To disable a policy, use the `tscli backup-policy disable` command.

Doing more with backup

The following table lists some additional backup commands you can use.

To	Command
List present backup policies.	<code>tscli backup-policy ls</code>
Show a backup policy.	<code>tscli backup-policy show <name></code>
Check the status of a policy.	<code>tscli backup-policy status <name></code>
Change an existing policy.	<code>tscli backup-policy update <name></code>
Disable or enable an existing policy.	<code>tscli backup-policy disable or enable</code>
Delete a policy	<code>tscli backup-policy delete <name></code>

Finally, you can time a `crontab` job with your periodic backup configuration to move a backup to longer term storage. Simply create a `crontab` job that moves the backup to a location outside of the `directory` defined in the periodic schedule.

About restore operations

Summary: Learn how to restore ThoughtSpot from a snapshot or backup.

When restoring to a running cluster where the ThoughtSpot software is not updated, we recommend that you use a snapshot. If you updated the cluster to a new release, if the configuration changed significantly, or when restoring to a different cluster, you must restore from a backup.

Restoring from backup require that you first delete the old cluster. Changes to a cluster that require restoring from a backup instead of a snapshot include:

- Removal of a node.
- Restoring to a different cluster from the one where the snapshot/backup was taken.
- Restoring to a cluster running a different release from the one where the snapshot/backup was taken.

You should never restore from a snapshot or backup yourself. To restore from a snapshot or backup, contact [ThoughtSpot Support](#) [See page 0].

Restoring a ThoughtSpot cloud deployment

For details on how to restore a ThoughtSpot cluster hosted in the cloud, see:

- [Backup and Restore with AWS](#) [See page 0]
- [Backup and Restore with GCP](#) [See page 0]

About data modeling

Summary: Modeling, tagging, and adding links between your data sources can make the data even easier to search.

Data modeling allows you to define metadata and other aspects of your data. For example, you can give data columns search friendly names or predefine how they can be explored and aggregated. Metadata include such information as **Column Names**, **Column Visibility**, **Column** and **Data Definition**, **Column Rank** and so forth.

When you load data, most of this data modeling metadata is set up for you automatically. However, since you know your data best, you can adjust the modeling settings to improve the experience for your users. After loading data, you can start searching your data without doing any data modeling, creating relationships, or tagging. However, since you know your data best, you can customize the modeling settings. Putting some thought into these will make the data even easier and more intuitive to search for your end users.

User interfaces for modeling data

Data modeling is a very lightweight process compared to what you may have experienced in other tools. It enables you to change some of the settings (or properties) of the data so that it becomes more searchable. You can configure the model [for an individual data table \[See page 406\]](#) or you can view and configure all the system data [using a modeling file \[See page 408\]](#). Editing the data model file requires that you have administrative privileges.

The model file contains a row for each column in your data set. It isn't unusual to have tens of thousands of rows in this file. This means that editing this file is equivalent to editing all the tables simultaneously. When you add new data to your system, this file expands to accommodate the new data columns you have added.

Both of these methods have the same effect: they improve search. Moreover, while they have different effects of scale, they use the same mechanisms to accomplish these effects.

Modeling topics

The following topics explain how to model your data:

- [Change a table's data model \[See page 406\]](#)

Explains how to make modeling settings for a table you've just loaded, or to make a quick change to existing settings.

- [Edit the system-wide data model \[See page 408\]](#)

Explains how to define a default data model to use for data system-wide.

- [Data modeling settings \[See page 413\]](#)

Explains the possible data model settings and their accepted values. These are the same for a table or the system.

- [Link tables using relationships \[See page 449\]](#)

Linked tables can be searched together or combined into a worksheet for easy searching.

Tables that have no relationship between their columns can not be combined in a single search.

- [About stickers \[See page 457\]](#)

You can create stickers to make it easier for people to find data sources and pinboards.

- [Manage experts or add an expert to your data \[See page 447\]](#)

Change a table's data model

Summary: You can adjust the data model for a newly loaded table.

To make modeling settings for a data source you've just loaded, or to make a quick change to existing settings, use the ThoughtSpot web interface. You can adjust the **Columns** settings from the data management listing.

You can change all the same data model settings here as in the model file. This method is easier and faster, unless you need to make many settings in bulk. In that case, [using the model file \[See page 408\]](#) is recommended.

About data sources

You can change the data modeling settings for base **Tables**, **Worksheets**, and **Views** [\[See page 494\]](#). Worksheets will inherit the data modeling settings from the tables upon which they are based. However, if you make further changes to a base table *after* you've created worksheets on it, the new data model changes will not propagate up. You must make any new data model changes directly to the worksheets (if you want them).

Change the data model for a data source

1. Click **Data** on the top navigation bar.
2. Click a data source you own or can edit.

The screenshot shows the ThoughtSpot Data interface. The top navigation bar includes 'Search +', 'Answers', 'Pinboards', 'SpotIQ', 'Data', and 'Admin'. Below the navigation is a search bar and a filter section with 'Tables' selected. The main area displays a list of tables with columns for 'Name', 'Source', 'Stickers', 'Materialize Status', 'Modified', and 'Author'. The table 'ThoughtSPORT_Product_Dimension' is highlighted with a yellow border.

Name	Source	Stickers	Materialize Status	Modified	Author
FoodDollarDataReal				47 minutes ago	P Plummer
ThoughtSPORT_Product_Dimension	Sports Goods			a month ago	AS Administrator Super-User
ThoughtSPORT_Retail_Sales_Fact	Sports Goods			a month ago	AS Administrator Super-User
MarketSpot_Vendor_Dimension	MarketSpot			a month ago	AS Administrator Super-User
ThoughtSPORT_Store_Dimension	Sports Goods			a month ago	AS Administrator Super-User
ThoughtSPORT_Customer_Dimension	Sports Goods			a month ago	AS Administrator Super-User

This brings up the **Columns** screen, where you can make your modeling settings.

3. Modify one or more column settings.

Descriptions of the possible settings are listed in [Data modeling settings \[See page 413\]](#).

4. Save your changes.

The screenshot shows the 'Columns' screen for the 'ThoughtSPORT_Product_Dimension' table. The top navigation bar includes 'Search +', 'Answers', 'Pinboards', 'SpotIQ', 'Data', and 'Admin'. Below the navigation is a 'Save Changes' button and a 'Load Data' button. The main area displays a table of column settings with columns for 'COLUMN NAME', 'DESCRIPTION', 'DATA TYPE', 'COLUMN TYPE', 'ADDITIVE', 'AGGREGATION', 'HIDDEN', 'SYNOMYS', and 'INDEX TYPE'. The table shows five rows of data.

COLUMN NAME	DESCRIPTION	DATA TYPE	COLUMN TYPE	ADDITIVE	AGGREGATION	HIDDEN	SYNOMYS	INDEX TYPE
Product_Key	Click to edit	INT32	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DONT_INDEX
Product_Name	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	product	DEFAULT
SKU_Number	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Department_Desc..	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT
Category	Click to edit	VARCHAR	ATTRIBUTE	<input checked="" type="radio"/> NO	NONE	<input checked="" type="radio"/> NO	Click to edit	DEFAULT

(showing rows 1-5 of 5)

5. To check your changes, use the **SEARCH** page to search for across the changed data.

Related information

- [Data modeling settings \[See page 413\]](#)
- [Edit the system-wide data model \[See page 408\]](#)
- [Understand data sources \[See page 0\]](#)

Edit the system-wide data model

Summary: Edit the modeling file to edit your data settings.

When you load data, ThoughtSpot uses defaults for data modeling metadata. You can change these defaults using the data modeling file if you have access to the **Data > Settings > Business Data Model** page. Editing this file allows you to view and edit all the system data columns. When you (or your users) add new data to your system, this file changes as it expands to accommodate new data columns.

Tip: If you just want to change a subset of your data, use the [Change a table's data model \[See page 406\]](#) instead.

Overview of the modeling process

The data formats you use in your system are controlled by the modeling file, an Excel file. To make these changes you download the model file, change the model, and upload your changes back into the system.

In each row of the modeling file, all the data properties corresponding to a column from your data are listed. You can modify many of these properties by typing in the new value. Remember these important guidelines when editing the model file:

- Do not modify any value in a column which contains **DoNotModify** in the field under the column heading.
- Make sure to keep the file in the same format as it had when you downloaded it.

The model file contains a row for each column in your data set. It isn't unusual to have tens of thousands of rows in this file. You can change all or a subset of rows. You can edit the file to leave the heading rows and only those rows you want to change. This can make the file more convenient to work with.

The model file must be saved as UTF-8 encoded. If your model file includes multi-byte characters, make sure you save it in the correct format or you won't be able to upload it after making your changes.

Download the model file

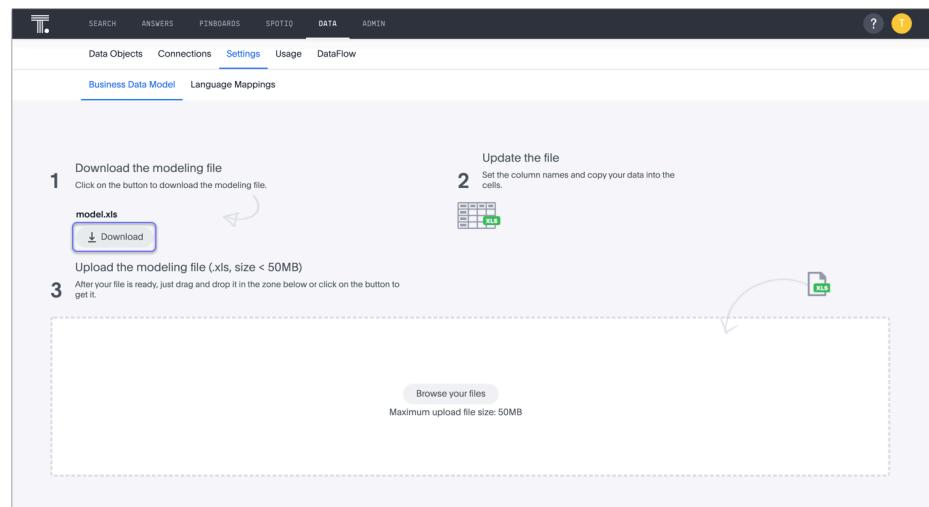
Before you can make changes to the model file, you need to download it. Then, you edit it using Microsoft Excel, vi/vim, or a similar text editing tool.

To obtain the model file:

1. Log in to ThoughtSpot from a browser as an Administrator user.
2. Click the **Data** tab in the top navigation bar.



3. Click **Settings**, then click **Business Data Model**.
4. Click **Download**.

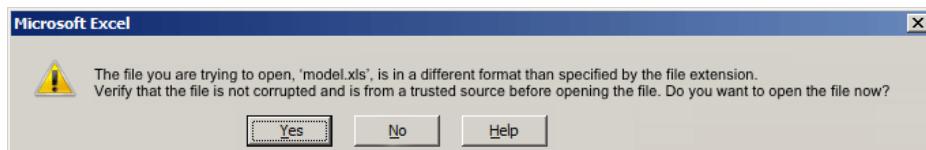


Edit the file and change the settings

You can make changes to the settings using this procedure. To see a list of the changes you can make, see [Data modeling settings \[See page 413\]](#). You can edit any of the values in the model file, except for those where the words **DoNotModify** appear under the column header. To make changes in the model file:

1. Open the model file you downloaded (`model.xls`) in Excel, vi/vim, or a text editor.

If you are using Excel, you may see a warning message.



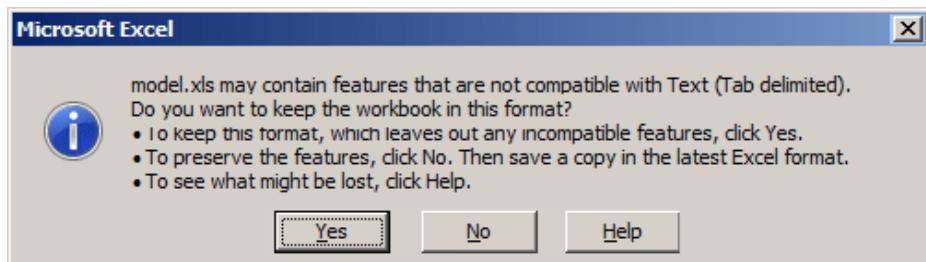
Click `YES` to proceed.

2. Find the column you want to modify.

Descriptions of the meanings of the columns are listed in [Data modeling settings \[See page 413\]](#).

3. Select the value you want to change.
4. Type in the new value.
5. After making all your changes, save the model file.

If you are using Excel, you will see a message. Click `YES` to save the file.



The model file must be saved as UTF-8 encoded. If your model file includes multi-byte characters, edit the file using vi or vim to ensure the file is saved in the correct format. Otherwise, you won't be able to upload it after making your edits.

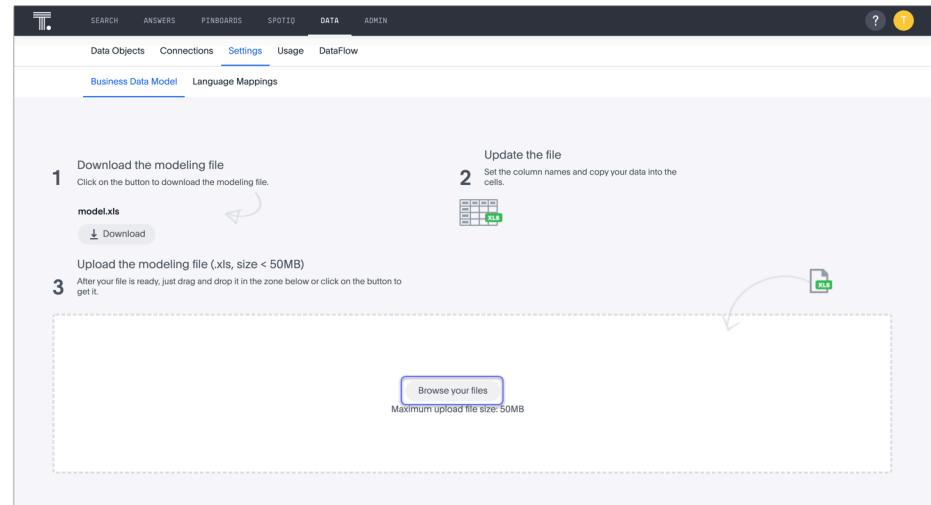
Upload the edited file

After you have made changes to the modeling file, you must upload it back to ThoughtSpot before the changes will take effect. To upload the model file:

1. Log in to ThoughtSpot from a browser as an Administrator user.
2. Click **Data**, on the top navigation bar.



3. Click **Settings**, then click **Business Data Model**.
4. Click **Browse your files** to upload the model.xls file, or drag and drop it in the zone.



If you receive an error message upon uploading the file, check that it does not include any multi-byte characters (for example, Japanese or other multi-byte language characters). If it does, you must download the file again and make your edits using vi or vim.

If you choose to remove all the rows you have not changed from the model file before uploading it. If you upload a model file that includes only the changed rows, you won't lose any of the pre-existing model file settings. This is a good option if your model file is causing an error on upload, but you aren't sure where in the model file the problem is.

As soon as the file is uploaded, ThoughtSpot performs any necessary re-indexing for you automatically. Your new settings will be reflected within a few minutes.

Related information

- [Data modeling settings \[See page 413\]](#)
- [Change a table's data model \[See page 406\]](#)

Overview of data modeling settings

Summary: Learn about data modeling settings.

You can change data modeling settings in two ways, both of which change the model. If you want to make a few small changes, you should [make them in the ThoughtSpot application \[See page 406\]](#). If you want to make many changes [you should edit the modeling file \[See page 408\]](#). Whether you are changing data modeling settings using the modeling file or the Web interface, the settings and their accepted values are the same.

Modeling settings

The following index includes mutable data modeling settings that you can apply to columns, both tables and worksheets.

Setting	Description
Column Name [See page 0]	Sets the name of the column to be used in searches.
Description [See page 417]	Adds a text description of what the column contains.
Data Type [See page 187]	Read only. Shows the column's data type.
Column Type [See page 418]	Sets the type of column, either ATTRIBUTE or MEASURE .
Additive [See page 420]	Controls the type of aggregations that will be available for a column.
Aggregation [See page 420]	Sets the default aggregation type for a column.

Setting	Description
Hidden [See page 424]	Sets the visibility of a column.
Synonyms [See page 424]	Adds synonyms that can be used in the search bar to refer to a column.
SpotIQ Preference [See page 426]	Excludes specified columns from SpotIQ analyses. By Default, all columns are included in SpotIQ.
Index Type [See page 428]	Sets the type of index that will be created for a column.
Geo Config [See page 434]	Enables a column to be used in GeoMap visualizations.
Index Priority [See page 428]	Changes the priority of a column in search suggestions.
Format Pattern [See page 438]	Specifies the format to use for numeric values or dates that show in the column.
Currency Type [See page 442]	Specifies the format of currencies in a column.
Attribution Dimension [See page 444]	Only applies to tables that join over a Chasm Trap [See page 208]. Designates whether the tables depend on this column for attribution. You cannot change the attribution dimension in the modeling file. You can only configure it on a table-by-table basis.
Calendar Type [See page 94]	Specifies what type of calendar a date type column uses. It can be Gregorian calendar (default), a fiscal calendar, or any custom calendar.
Entity Category [See page 577]	Specifies how to categorize the data in the column: person, place, time, and so on. Important for configuring SearchIQ Beta .
SearchIQ Enabled	Indicates if the column is enabled for SearchIQ Beta .

Data modeling best practices

As a best practice, make any data modeling settings in the table when you will be creating multiple worksheets that use that table. This way, you won't have to make the same settings in each worksheet. The settings will be inherited when you create worksheets that uses columns from the table.

If you have settings that only apply in the context of a particular worksheet, make those settings in the worksheet rather than in the underlying table(s).

Note that if you make your settings at the table level, and then create a worksheet that uses columns from the table, the settings are inherited from the table at the point in time that the worksheet is created. If you then go back and change the settings at the table level, your changes will not be reflected in the worksheet.

If you want the worksheet to have the changes you made at the table level, you must drop those columns from the worksheet and re-add them. Then save the worksheet. At this point, the new settings will be used in the worksheet. Note that any saved answers or pinboards based on the worksheet may display differently because of your changes. For example, if you've changed the GeoMap setting from "None" to "Country", you will now see a map where before you might have seen a table.

Related information

- [Model the data for searching \[See page 404\]](#)
- [Add a geographical data setting for a column \[See page 434\]](#)

Set column name, description, and type

Summary: Modeling includes setting basic information for a data column such as its name, description, and type.

When you model your data, you set basic information for data columns, such as a column's **NAME**, **DESCRIPTION**, and **TYPE**. All of these can influence how a user experiences your data. For example, the **DESCRIPTION** appears as a "tip" when a user hovers over a column. It helps users understand where the data they are searching comes from.

Note: When you change column information such as column names in the UI or in the model file, you do not change this information on the physical table in your database. You only change your metadata. This is why you can have column names with spaces in them in the UI, since physical column names do not support spaces.

Data modeling from the UI

To model your data columns from the ThoughtSpot UI, follow these steps:

1. Navigate to the **Data** tab from the top navigation bar.
2. Click on the data source that you would like to update.
3. Click **Edit Worksheet** in the top right corner of the screen.

COLUMN NAME	DESCRIPTION	DATA TYPE	COLUMN TYPE	ADDITIVE	AGGREGATION	HIDDEN	SYNONYMS	SPOTIQ PREFERENCE
Sales Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Sales Amt, Reve...	DEFAULT
Cost Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Click to edit	DEFAULT
Store Name	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT
Product	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT

4. Update your data columns.

Change the column name

The **Column Name (UI)/ ColumnName** (model file) parameter is the name users type to add a column to their search. Change these column names in ThoughtSpot to make them more meaningful to users.

The model file contains a row for each column in your data set. It isn't unusual to have tens of thousand of rows in this file. You can change all or a subset of rows. You can edit the file to leave the heading rows and only those rows you want to change. This can make the file more convenient to work with.

The default column name is the name you gave the column when you defined the table in the database or imported the CSV file from the browser.

To update the column name, follow these steps:

1. Find the column name you want to change, either on the ThoughtSpot application or in the model file.
2. Type in the new column name.
3. Save your changes.

Change column description

The **Description (UI)/ ColumnDescription** (model file) parameter is an optional description for a column. You can provide a description for a specific column, to provide additional information for users about the data it contains. When a user hovers over the column, a tooltip will show this description.

To create a column description:

1. Find the column description you want to change, either on the ThoughtSpot application or in the model file.
2. Enter a new description.
3. Save your changes.

Change column type

The **Column Type** (UI)/ **ColumnType** (model file) parameter describes the kind of data a column stores.

This is set automatically upon defining the table, but in some cases, you may want to change the type.

There are two types of columns:

- **ATTRIBUTE** contains a characteristic or trait associated with your data, such as `name`, `address`, or `id number`.
- **MEASURE** contains a numeric value that can be compared in a meaningful way using math, such as a count or measurement, like `sales`.

When you create a new table, the default column type is set according to the **Data Type** defined for each column. By default, columns with numeric data types (`FLOAT`, `DOUBLE`, `INT`, or `BIGINT`) are assigned the type `MEASURE`. Columns with `VARCHAR`, `BOOL`, or date/time data types are assigned the type `ATTRIBUTE`.

Usually the default setting for column type works fine. But occasionally, you must change a `MEASURE` to an `ATTRIBUTE`. Examples of numeric values for which mathematical operations are not meaningful include:

- ID numbers
- Key values that are primarily used for joining tables
- Product number or SKU
- Sports team member jersey number
- Year, when separate from a date (e.g. 1999, 2000)
- Zipcodes

These values are numbers, but you would not do math on them. For example, it is not meaningful to add two zipcodes together.

To change the column type:

1. Find the column type you want to change, either on the ThoughtSpot application, or in the model file.
2. Change it to either `MEASURE` or `ATTRIBUTE`.
3. Save your changes.

Related information

- [Model the data for searching \[See page 404\]](#)
- [Hide column or define a column synonym \[See page 424\]](#)

Set ADDITIVE or AGGREGATION

Summary: You can allow aggregation on MEASURE columns and some ATTRIBUTE columns.

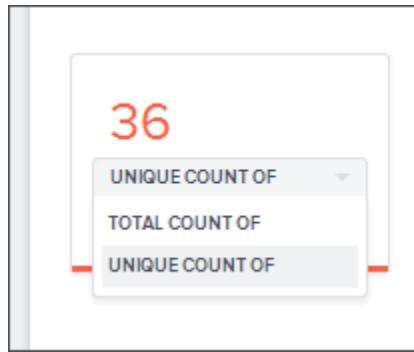
Aggregation is grouping many units or parts into a new value. In data, aggregation gathers multiple input values and calculates a summary value from them. You then use this aggregated value to do an analysis.

Every summary value results from a data aggregation function. An example aggregation function would be average or minimum. You can control how aggregation works in your data.

Making an ATTRIBUTE column ADDITIVE

Your data may contain a column with a numeric data type that you have defined as an ATTRIBUTE rather than a MEASURE. For example, you may have ATTRIBUTE column with an INTEGER data type that represents age. Typically, these columns have an ADDITIVE setting of NO. Within a search result that contains data from this column, the options for each column on the left side of the screen includes:

- UNIQUE COUNT OF
- TOTAL COUNT OF

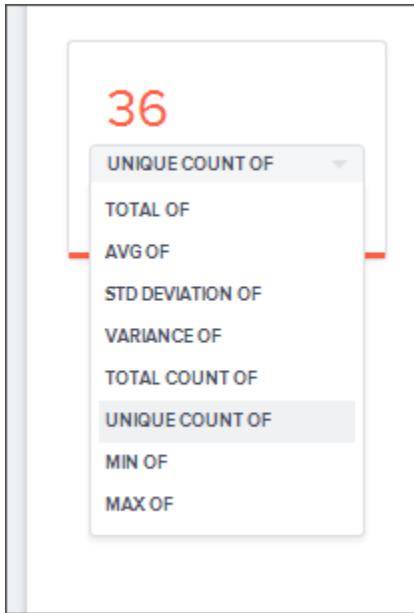


To display extended aggregate view options, you must set ADDITIVE to YES on these ATTRIBUTE columns. This option is only possible on columns that have a numeric data type (FLOAT , DOUBLE or INTEGER) or a date data type (DATE , DATETIME , TIMESTAMP , or TIME). After you make this change, these additional view options area-charts offered:

- TOTAL OF

Set **ADDITIVE** or **AGGREGATION**

- **AVG OF**
- **STD DEVIATION OF**
- **VARIANCE OF**
- **TOTAL COUNT OF**
- **UNIQUE COUNT OF**
- **MIN OF**
- **MAX OF**



To change this setting:

1. Find the column whose **ADDITIVE** setting you want to change
2. Select the **ADDITIVE** toggle.
3. Change the value to one of these:
 - YES or NO, if using the Web interface.
 - TRUE or FALSE, if using the model file.
4. Save your changes.

Change Aggregation

Both `MEASURE` columns and `ATTRIBUTE` columns support **AGGREGATION** operations. To aggregate a column without having to enter the aggregation type explicitly in your searches every time, you can set a default **Aggregation** for that column. Setting this default can make combining data more intuitive and faster.

`ATTRIBUTE` columns have **AGGREGATION(UI)/AggregationType** (model file) values with default aggregate type of **NONE**. You can change **AGGREGATION** to one of the supported aggregation types. To extend the available aggregation actions, set **ADDITIVE** on these columns to `YES` (`TRUE`).

Aggregate type	Description
NONE	Does no aggregation. This is the default for <code>ATTRIBUTE</code> type columns.
SUM	Adds the values together and returns the total. This is the default for <code>MEASURE</code> type columns.
AVERAGE	Calculates the average of all the values.
MIN	Calculates the minimum value.
MAX	Calculates the maximum value.
STD_DEVIATION	Calculates the standard deviation of all the values.
VARIANCE	Calculates the variance of all the values.
COUNT	Calculates the total number of values.
COUNT_DISTINCT	Calculates the total number of distinct values.

Keep in mind that not all `MEASURE` data should be aggregated. Consider a table containing data about athletes on a sports team. The data contains some numerical values, including points scored, salaries, and jersey numbers for each of the players. Because jersey number is an `INTEGER`, it would become a column of type `MEASURE` (not `ATTRIBUTE`). So it will aggregate, by default. But you may want to make its aggregation type **NONE** instead. This ensures that search results that include jersey number will not attempt to compare or aggregate those values in a way that is not meaningful.

To set this value.

1. Find the column whose default aggregation type you want to change
2. Select its **Aggregation**. If using the modeling file, use the **AggregationType** setting.
3. Select the new default aggregation type.
4. Save your changes.

Related information

[Model the data for searching \[See page 404\]](#)

Hide a column or define a synonym

Summary: Hide a column from users or make it easier to find by assigning a synonym.

You can hide columns from users in ThoughtSpot without dropping them from the database. It is common to hide a column when its data contains identifier columns that are used to join tables, but which do not have any meaning to users.

Alternatively, rather than hiding a column, you can make it easier to find by creating synonyms for it. This is helpful, for example, when different departments refer to the data using different terminology.

Hide a column

As the number of columns in the dataset increases, the search experience requires more effort. Users have to navigate through larger numbers of columns to choose the correct one. There might also be some columns in the dataset that you don't want to expose to the users.

Change the **HIDDEN (UI)/Hide** (model file) setting to hide a column. By default, all columns in a data source appear in ThoughtSpot. To hide these columns, set the **HIDDEN** setting to `YES`.

1. Find the **HIDDEN (UI)/Hide** (model file) setting for a column.
2. Set its value to `YES`.
3. Save your changes.

Create synonyms for a column

When users search a data source, they might try typing different words to try to retrieve a particular column. This could be due to different groups in your organization using different terms for the same data. Or maybe your users just intuitively use different words when searching for that item. Using synonyms allows them to access the data even if the term they choose isn't the same as the actual column name.

You can set column synonyms for columns in tables, user imported data, and worksheets. The returned table or chart uses the *actual column name*, but the search bar reflects the term the user typed in (the synonym).

To create a synonym for a column:

1. Find the column for which you want to add synonyms.
2. Select its **Synonyms**.
3. Enter a comma-separated list of the synonyms.

If a synonym is more than one word, enclose it in double quotes. If you are using the Web interface, you would type:

```
profit,"gross profit"
```

If you are using the model file, the list of synonyms must be enclosed in square brackets:

```
[profit, "gross profit"]
```

4. Save your changes.

Related information

[Model the data for searching \[See page 404\]](#)

Set columns to exclude from SpotIQ analyses

Summary: You can specify columns to exclude from SpotIQ analyses.

SpotIQ [See page 0] is a ThoughtSpot feature that provides users with insights about their data by automatically surfacing interesting characteristics (trends, correlations, outliers, and so on).

If you have access to tables, worksheets, and views for data modeling purposes, you can specify columns to exclude from SpotIQ analyses. By default, all columns are *included* in SpotIQ analyses.

Exclude columns from SpotIQ analyses

To specify columns to exclude from SpotIQ analyses:

1. Click **Data** in the top menu, and choose **Tables**, **Worksheets**, or **Views** [See page 494].
2. Click the name of your data source.
3. On the **Columns** tab, find the COLUMN NAMES you want to exclude from SpotIQ analyses, and scroll to the right to find **SPOTIQ PREFERENCE**.
4. Use the drop-down menu to set the **SPOTIQ PREFERENCE** to **EXCLUDE** for each column you want to exclude.
5. Click **SAVE CHANGES** in the upper right.

Include columns in SpotIQ analyses

By default, all columns are included in SpotIQ analyses.

If you have previously set some columns to EXCLUDE and you want to re-set these to be included, do the following.

1. Click **Data** in the top menu, and choose **Tables**, **Worksheets**, or **Views** [See page 494].

2. Click the name of your data source.
3. On the **Columns** tab, find the COLUMN NAMES you want to set back to include in SpotIQ analyses, and scroll to the right to find **SPOTIQ PREFERENCE**.
4. Use the drop-down menu to set the **SPOTIQ PREFERENCE** to **DEFAULT** for each column you want to include.
5. Click **SAVE CHANGES** in the upper right.

Related information

- [SpotIQ tutorial \[See page 0\]](#)
- [Overview of data modeling settings \[See page 413\]](#)

Manage suggestion indexing

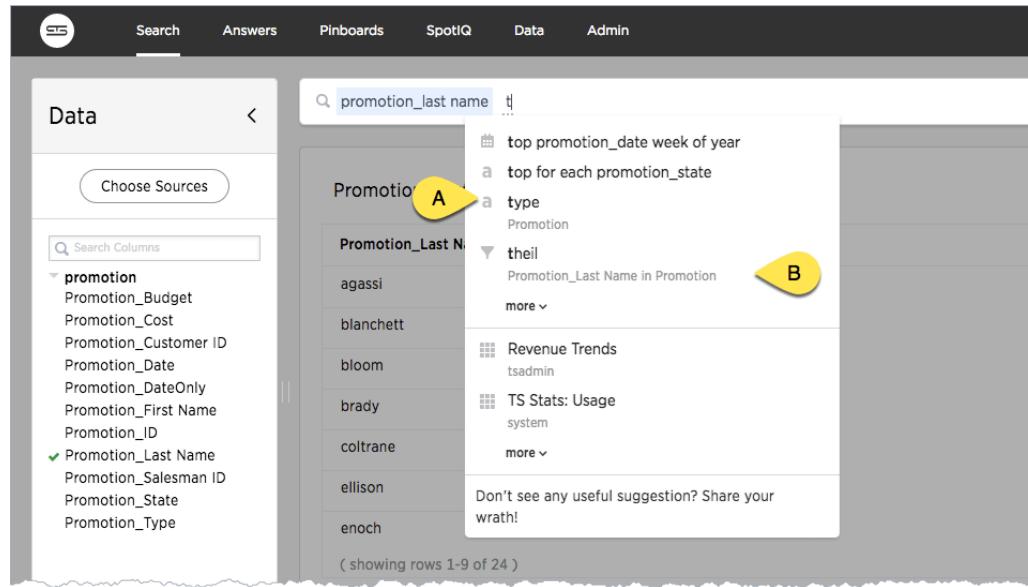
Summary: ThoughtSpot dynamically indexes Search bar suggestions for column names and values.

When a user searches in the **Search** bar, ThoughtSpot supplies the user with suggestions for column names and their column values. The **COLUMN NAME** and any **SYNONYMS** appear in **Search** suggestions. A column's **INDEX TYPE** controls whether and how ThoughtSpot suggests column values.

Additionally, ThoughtSpot uses a column's **INDEX PRIORITY** value to determine where to rank a column's name and values in the search suggestions. These values impact the dynamically calculated *usage based ranking (UBR)*,

Example of Search suggestion behavior

The following example illustrates how searching for `promotion_last_name t` causes the system to suggest several ways of completing the `t` in the search:



The system is suggesting the synonym `type` (callout A) for a column in the `Promotion` table. It is also suggesting a value of `theil` (callout B) for the `Promotion_Last Name` column. If you look in the **Data > Tables** page, you can see that there is a `type` synonym for the `Promotion_Type` column which is using default indexing.

COLUMN NAME	AGGREGATION	HIDDEN	SYNOMYS	INDEX TYPE	GEO C...
Promotion_Type	NONE	<input checked="" type="checkbox"/> NO	Type	DEFAULT	None
Promotion_Date	NONE	<input checked="" type="checkbox"/> NO	Click to edit	DEFAULT	None
Promotion_Sales..	NONE	<input checked="" type="checkbox"/> NO	Click to edit	DEFAULT	None

Managing search suggestions through **INDEX TYPE** and **INDEX PRIORITY** is important. Properly configured suggestions can decrease “noise” in the suggestion list. Increasing the visibility of important columns is helpful for new or intermittent ThoughtSpot users.

Understand the default indexing behavior

ThoughtSpot has a system default **INDEX TYPE** behavior for search suggestions. This system default is configured on your cluster and applies to all worksheets and tables. You can override this default behavior on a per-column basis.

The system behavior when the **INDEX TYPE** is **DEFAULT** is as follows:

- With two exceptions, the system indexes all columns using their **COLUMN NAME** value. The exceptions are columns with **COLUMN TYPE** of `MEASURE` and columns with **DATA TYPE** of `DATE`.
- Columns that contain data values with large amount of free-form strings, that is, a length is greater than 50 words, are indexed as `PREFIX_ONLY` by default.

Warning: If a column has a very large free text values, ThoughtSpot recommends you keep **DEFAULT** or set **DONT_INDEX**. Other settings indexing on these values may generate confusing suggestions.

- Short strings (like a `firstname` column) are indexed using `PREFIX_AND_SUBSTRING` by default, which indexes both prefix and substrings.
- If a column is using has a *cardinality* – the number of unique column values – greater than 10 million, it is not indexed.

If a column's **INDEX TYPE** is *not* **DEFAULT** and the column's cardinality is greater than 30 million, ThoughtSpot does not index the column.

High cardinality and performance

A column's cardinality can impact indexing. If you have a column with a very high cardinality and a very high number of rows, indexing these values can impact your ThoughtSpot performance. ThoughtSpot Support recommends you turn off indexing of primary key columns on extremely large tables (> 10 million rows) in your cluster.

High cardinality is relative to other considerations. In some cases, columns with fewer than 10 million rows but with columns containing long strings can cause performance problems with memory. If you have concerns or questions, your ThoughtSpot Customer Success Engineer can help you determine appropriate cardinality thresholds for your ThoughtSpot installation.

Configure your own cluster defaults

If you need to, you can work with ThoughtSpot Support or your Customer Success Engineer to configure new cluster defaults.

Override the system default on a column

You can change a column's **INDEX TYPE** in the **Data > Tables > Columns** page or in the **Index** value in the modeling file.

The values you can set for **INDEX TYPE** are:

Index type	Description
------------	-------------

DEFAULT	The default behavior applies to all ATTRIBUTE columns that are not DATE types. PREFIX_AND_SUBSTRING for short values and PREFIX_ONLY for long values and free-form text.
DONT_INDEX	Prevents indexing on the column values. The column doesn't appear in search suggestions.
PREFIX_AND_SUBSTRING	Allows full indexing such that prefix and sub-string search both work for the column values.
PREFIX_AND_WORD_SUBSTRING	Allows indexing such that only prefix search works for each word of a multi-word string, for the column values.
PREFIX_ONLY	Allows indexing such that only prefix search works for the column values.

Consider a column in which there are four values ThoughtSpot , Thought , Spot and ThoughtSpot . If you search for sp , depending on the setting for indexing, the column value search result suggestions will vary:

Index field value	Search bar suggestions
DEFAULT	ThoughtSpot , Spot and ThoughtSpot
DONT_INDEX	No suggestions.
PREFIX_AND_SUBSTRING	ThoughtSpot , Spot and ThoughtSpot
PREFIX_ONLY	Spot
PREFIX_AND_WORD_SUBSTRING	Spot and ThoughtSpot

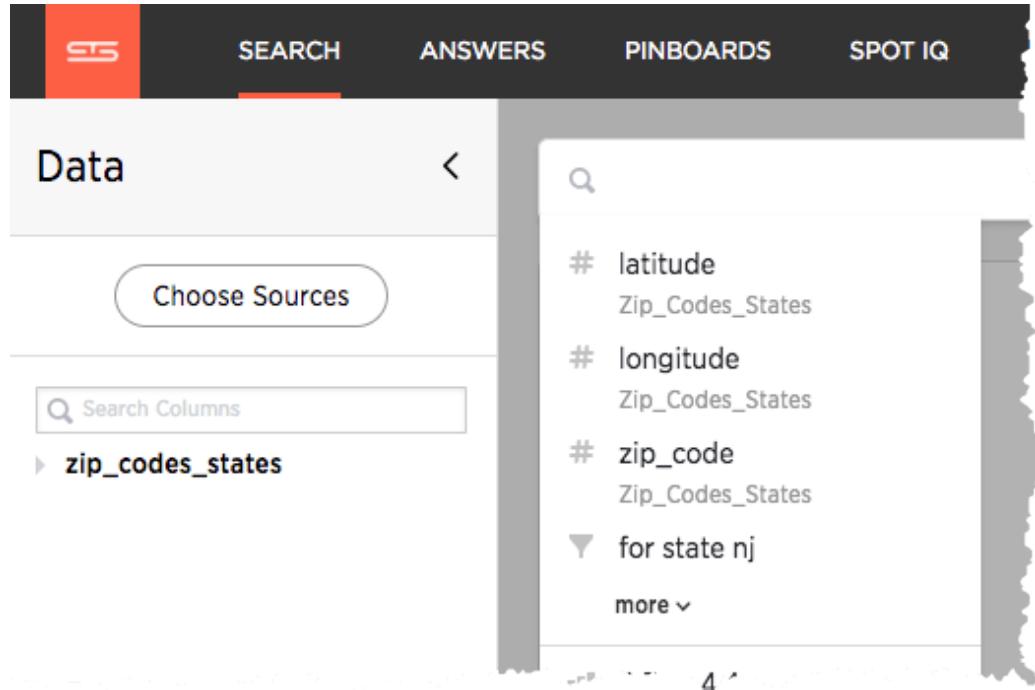
To change a value in the application UI:

1. Open a worksheet or table from the **Data** page.
2. Find the column whose index type you want to modify.
3. Set its **INDEX TYPE**.
4. Save your changes.

If you are using the model file, locate the **Index** cell, and enter the **INDEX TYPE** you want to use.

Change a column's suggestion priority

A column's **INDEX PRIORITY** determines the order or rank in which it and its values appear in the search dropdown.



The screenshot shows the ThoughtSpot interface with the following elements:

- Top Navigation Bar:** SEARCH, ANSWERS, PINBOARDS, SPOT IQ.
- Main Area:** A title "Data" with a back arrow, a "Choose Sources" button, and a search bar labeled "Search Columns".
- Search Dropdown:** A list of suggestions starting with "# latitude Zip_Codes_States", "# longitude Zip_Codes_States", "# zip_code Zip_Codes_States", and "for state nj". There is also a "more" link at the bottom of the list.

By default, the **INDEX PRIORITY** value is set to `1` for all columns. You can push a column up in the order (increase the rank) by increasing its **INDEX PRIORITY** value. A higher value (like `2`) will cause the corresponding column and its values to appear higher up in the search dropdown than columns with lower value (like `1`).

COLUMN NAME	CONFIG	INDEX PRIORITY
zip_code		1
latitude		10
longitude		1
city		1
state		1
county		1

You should only use numbers between 1-10 in the **INDEX PRIORITY** field. Use a value between 8-10 for important columns to improve their search ranking. Use 1-3 for low priority columns.

To change a value in the application UI:

1. Open a worksheet or table from the **Data** page.
2. Find the column whose index type you want to modify.
3. Change the **INDEX PRIORITY** to a number between 1 and 10.
4. Save your changes.

If you are using the model file, locate the **Index** cell, and enter the priority you want to use.

Related information

- [Model the data for searching \[See page 404\]](#)
- [Usage based rankings \(UBR\) \[See page 0\]](#).

Add a geographical data setting

Summary: Learn how to model your geographical data.

Certain attribute columns that contain location data can be used to create GeoMaps. ThoughtSpot supports Latitude, Longitude, Zip Code, US States, US Counties, Countries, and select international sub-nation regions.

You can designate a column as `Geo` by editing the **GEO CONFIG** column in the table **Columns** page.

You cannot edit the geo configuration column information in the `model.xls` file.

Guidelines for geographic columns

Columns that can be designated as `Geo` columns need to contain text (`VARCHAR`) data unless they contain latitude/longitude data. Latitude and longitude columns can contain numeric data (`DOUBLE`) or text.

If you import `geo` columns as numeric values, the data type defaults to `BIGINT`. ThoughtSpot recommends that you import `geo` columns, such as zipcodes as text values. The `column type` defaults to `measure` when you import numeric values. In ThoughtSpot, a `measure` is a numeric value that you can use in mathematical formulas. For geo data, you **must** change the `column type` to `attribute` and specify `additive` as `no`.

If you are using a column with the data type `DOUBLE` for latitude and longitude, change the following settings for those columns:

- set **Column Type** to `ATTRIBUTE`
- set **Additive** to `NO`
- set **Aggregation Type** to `NONE`

For information on these settings, see Set ADDITIVE or AGGREGATION [See page 420].

As a best practice, you should make the following changes to your geographical data **before** importing it to ThoughtSpot:

1. Specify `geo` columns as text values, and not numeric.

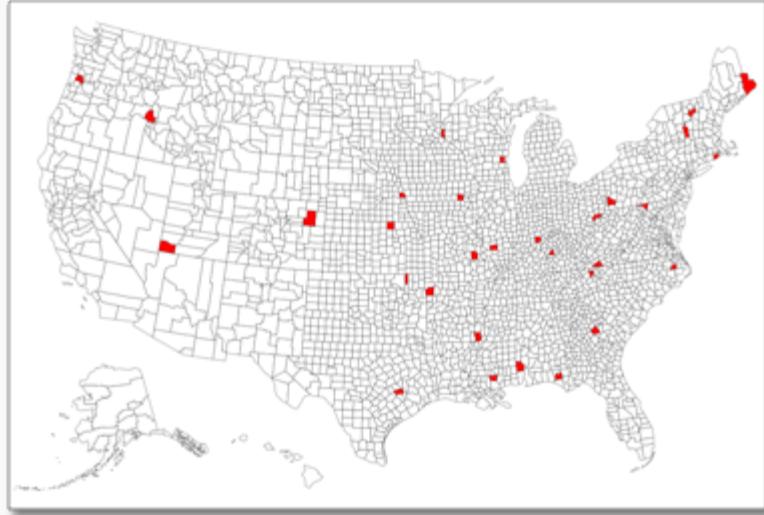
2. Make sure your data is specific, and does not contain duplicates.

For example, you may be collecting data for locations in Washington County, Alabama, and also for locations in Washington County, Wisconsin.

According to the 2010 Census, there are 31 distinct Washington Counties in the United States.

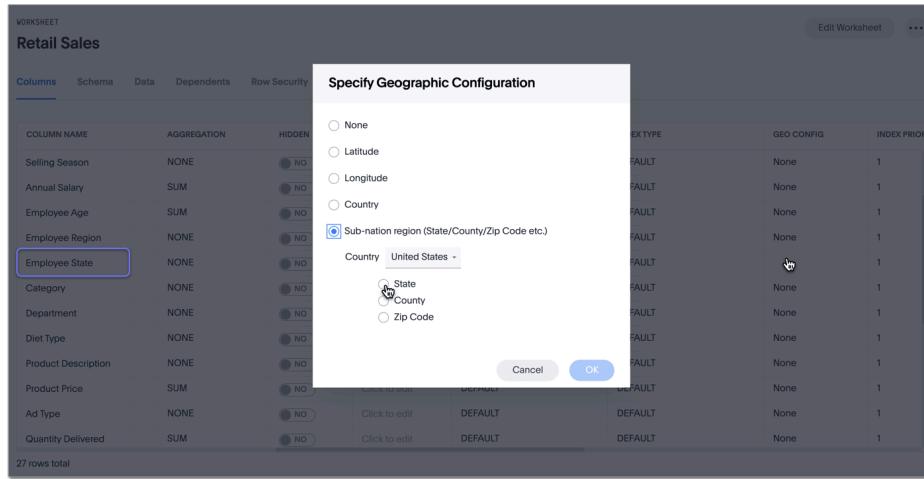
To account for multiple different geographical entities with the same name, you must be specific within your data.

A value of *Washington County* without a state attached to it may result in incorrect or incomplete geo maps.



How to edit geographic columns

1. Find the **GEO CONFIG** for the column that contains the geographical data.
2. Select the column to display the **Specify Geographic Configuration** dialog.



3. Change the value to the appropriate **GEO CONFIG**, depending on the kind of geo data the column contains.

If your data includes latitude and/or longitude columns that are stored as a numeric data type (`DOUBLE`), make these changes for those columns:

- a. Change the **Type** to `ATTRIBUTE`.
 - b. Change **ADDITIVE** to `NO / FALSE`.
4. Save your changes.

Supported geo maps

For a complete list of supported geo maps in ThoughtSpot, refer to: [Geo Map Reference \[See page 0\]](#)

Related information

[Model the data for searching \[See page 404\]](#)

Set number, date, and currency formats

Summary: Learn how to set key formats for column values.

You can set number, date, and currency display formats. These formats define how these value types display in tables and charts.

Number formats

You can set a format for how numbers are displayed in tables and charts. For example, you can display numbers with a different number of digits after the decimal point, based on the data modeling setting

Format Pattern. You can use any of the supported number formats for delimiters and number of digits to show using [Java Decimal Notation](http://docs.oracle.com/javase/7/docs/api/java/text/DecimalFormat.html) (<http://docs.oracle.com/javase/7/docs/api/java/text/DecimalFormat.html>). Currency symbols are not supported.

The system has default values which are:

`#,##` For integer data types `INT` and `BIGINT`. As you can see, these can only contain numbers, alpha characters are not permitted.

`#,###.00` for decimal data types `DOUBLE` and `FLOAT`.

These are some examples of formats you can use:

Stored Value	Format Pattern	Display Value
12345.6789	<code>#,##0.##</code>	12,345.68
12345.6789	<code>#,##0.###</code>	12,345.679
12345.6789	<code>#,##0.00000</code>	12,345.68
12345.6789	<code>#,##0</code>	12,345
12345.6789	<code>#,##0.00</code>	12,345.68
12345	<code>#,##0.##</code>	12,345

Stored Value	Format Pattern	Display Value
12345	,##0.00	12,345.00

You can change the date format used to display a column's values [for a single table \[See page 406\]](#) or, by editing the data model, for [the entire ThoughtSpot instance \[See page 408\]](#). Editing the data model file requires that you have administrative privileges.

1. Decide if the change is for a table or the entire instance.
2. Find the **Format Pattern** for the column.

This is either a column in a single table or a column in the data modeling file.

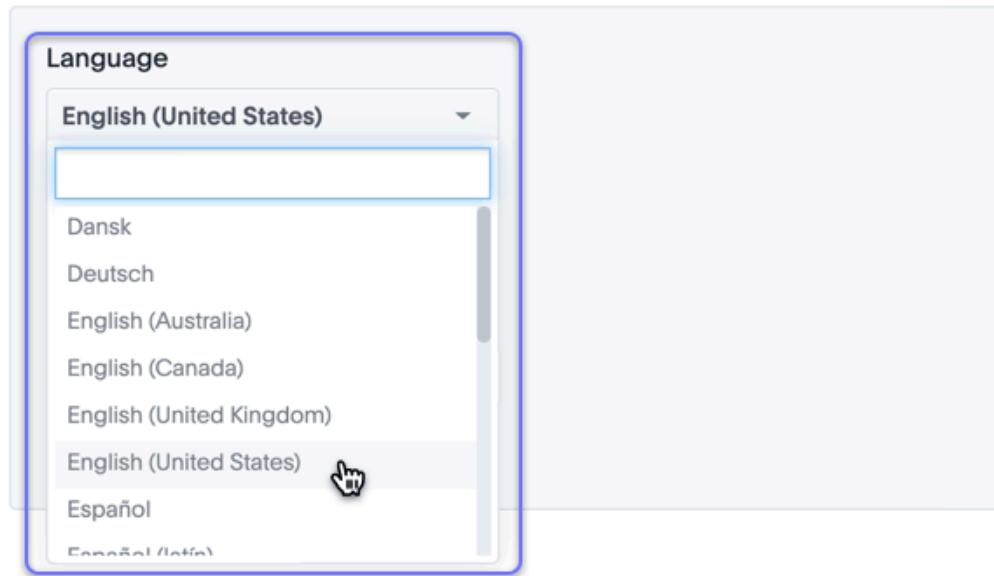
3. In the column, enter the format you want to use.
4. Save your changes.

If you are using a data-modeling file you must upload the new file to your installation.

Profile-based number formatting

Number formatting is set by default based on your ThoughtSpot profile's **Preferred locale** setting. You can set this value to accommodate your geographic locations.

Preferences



For example, if you are using ThoughtSpot in the US, the number formatting should look like this:

xxx,xxx.xx . And in Europe, it should look like this: xxx.xxxx,xx .

Date formats

Format Pattern (UI)/ Format Pattern (model file) formats for how dates are displayed in tables and charts. For example, you can display dates in a standard European or US format based on the data modeling setting **Format Pattern**. These are some examples of formats you can use:

Format mask	Description
YYYY or yyyy	four digit year such as 2017
YY or yy	last two digits of year such as 17
M	month with no leading zero 1 - 12
MM	Two digit month 01 - 12
MMM	Three letter month such as Jan
D	Day of year without a leading zero 0 - 365
DD	Day of year with up to one leading zero 01 - 365

Format mask	Description
DDD	Day of year with up to two leading zeroes 001 - 365
d	Day of month with no leading zero 1 - 31
dd	Two digit day of month 01 - 31
HH	Two digit 24 hour representation of hour 00 - 23
hh	Two digit 12 hour representation of hour 01 - 12
H	24 hour representation of hour with no leading zero 0 - 23
h	12 hour representation of hour with no leading zero 1 - 12
mm	Minutes 00 - 59
m	Minutes with no leading zero 0 - 59
ss	Seconds 00 - 59
s	Seconds with no leading zero 0 - 59
a	AM/PM indicator

Valid delimiters include most non-alphabet characters. This includes but is not limited to:

- / (forward slash)
- \ (backward slash)
- | (pipe symbol)
- : (colon)
- - (dash)
- _ (underscore)
- = (equal sign)

Examples of valid format masks you can produce for display are as follows:

- MM/dd/yyyy
- MMM
- DD/MM/yyyy
- MM/dd/yyyy HH:mm
- DD/MM/yyyy HH:mm

To change the date format used to display a column's values for a single table [See page 406] or, by editing the data model, for the entire ThoughtSpot instance [See page 408].

1. Decide if the change is for a table or the entire instance.

2. Find the **Format Pattern** for the column.

This is either a column in a single table or a column in the data modeling file.

3. In the column, enter the format you want to use.

4. Save your changes.

If you are using a data-modeling file you must upload the new file to your installation.

Set currency type

You can set a format for how currencies display in tables and charts when using the ThoughtSpot Data API or embedding. For example, you can display currencies in a standard European Euro or US Dollar format, based on the data modeling setting **Currency Type**.

You can change the currency format used to display a column's values for a single table [See page 406].

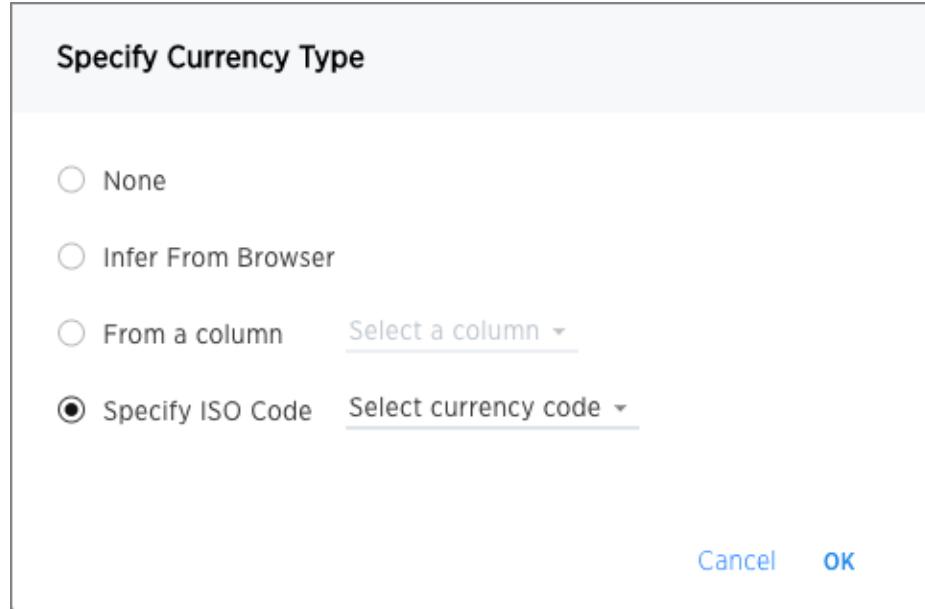
When you specify the currency type of your data in the **Columns** settings, your currency data will only display the correct format and currency code in the embedded use case. Currency specific symbols are available in the non-embedded use case as well, but they are not localized.

All users are treated as if they are in `en-US` locale unless they are in embed mode and their browser configuration tells ThoughtSpot that they are in some other locale. For example, `100 Polish Zloty` appears as `100 zł` to a user in Poland, but without localization enabled, it appears as `PLN 100`.

This subtle difference can be seen when you use the REST API. See the ThoughtSpot Application Integration Guide for more information on the API.

1. Find the **Currency Type** for the column whose display format you want to change.

2. Click it to open the **Specify Currency Type** menu.



3. Select one of the following ways you would like to change the format.

Option	Description
Infer From Browser	Your currency data will be modeled upon the locale of your browser setting.
From a column	Your currency data will be modeled upon the existing currency information in the selected column. This option is disabled if there is no VARCHAR column to choose from.
Specify ISO Code	Your currency data will be modeled upon your selection from the available currency code choices.

4. Click **Ok** to save your changes.

Related information

[Model the data for searching \[See page 404\]](#)

Change the Attribution Dimension

Summary: The Attribution Dimension setting applies only to tables that are related through a chasm trap. If your schema does not include these, you can ignore this setting.

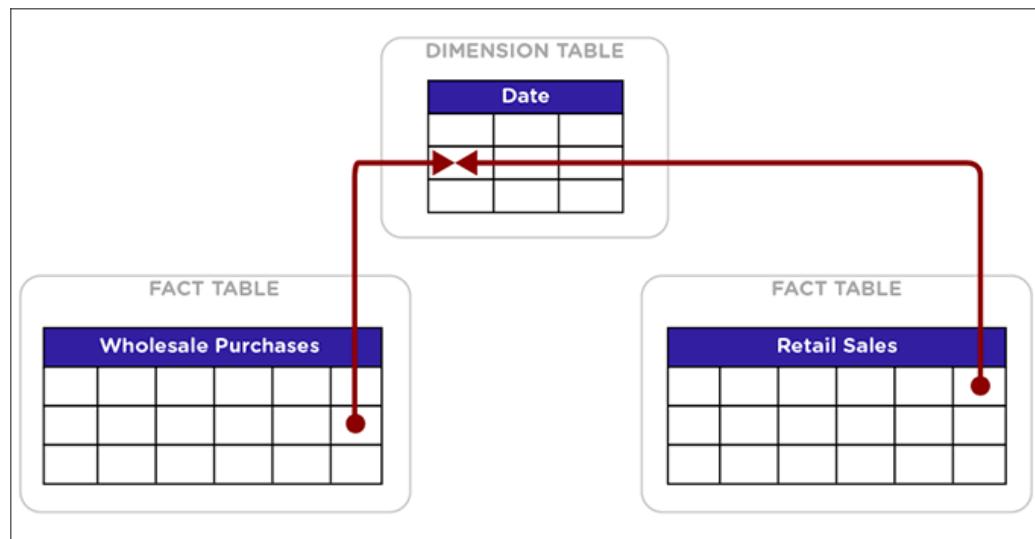
The **Attribution Dimension** setting only applies to tables that join over a Chasm Trap [See page 208]. By default, the attribution dimension setting will be set to `YES`, but you can override that by setting the column's attribution dimension property to `N0`, as described here.

Understand chasm traps and attribute dimension

In the classic chasm trap, two fact tables are related through a shared dimension table. When the two fact tables are joined, the shared column(s) in the dimension table are used to attribute rows in one fact table to match with rows in the other fact table.

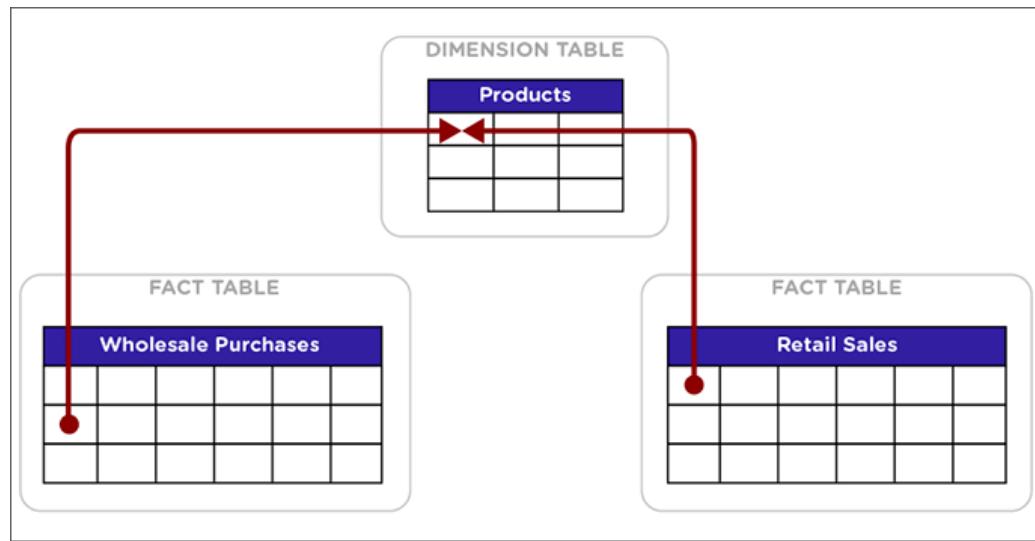
Usually, all goes well using this method. But sometimes an incorrect or illogical attribution can result. This can happen because the column chosen is not meaningful for performing this attribution. If you are seeing unexpected results in searches that include tables across a chasm trap, this setting is for you.

Below is an example of a column that is not an attribution dimension. Suppose you have two fact tables, Wholesale Purchases and Retail Sales, that share a common dimension Date.



In this example, the date column in the Date dimension should not be used for attribution, since unrelated rows in both of the fact tables could share the same row in the Date table. Why? Because if Sally bought oranges wholesale on April 25, 2005 and made a retail sale of apples on the same day, there is no logical relationship between those two events. Combining the two events using the date they share will not create any meaningful information.

If matching rows in two fact tables over a chasm trap depends on the values in a column contained in a dimension table, that column is known as an attribution dimension.



In this example, the Product ID column in the Products dimension table is an attribution dimension. For rows where the Product ID in the Wholesale Purchases and in the Retail Sales tables is a match, those rows are logically related in a meaningful way. They can be combined in charts and reports to produce a logical, expected outcome.

How to set attribute dimension

You cannot configure this setting in the model file. You can only configure it on a table-by-table basis. To designate a column as not being an attribution dimension (not producing any meaningful attribution across a chasm trap):

1. Navigate to the dimension table you would like to edit. Note that you do not need to set the Attribution Dimension for fact tables.
2. Find the column that is not an attribution dimension. Since the Attribution Dimension is a table-wide setting, you only need to turn it off on one column. As a best practice, use the

primary key.

3. Select its **Attribution Dimension**.
4. Set the value to **NO**.
5. Save your changes.

Additional resources

As you develop your expertise in data modeling and the Attribution Dimension setting, we recommend the following ThoughtSpot Community article:

- [What are Chasm Traps, and What is Attribution? \(<https://community.thoughtspot.com/customers/s/article/What-is-Attribution-and-Chasm-Traps?>\)](https://community.thoughtspot.com/customers/s/article/What-is-Attribution-and-Chasm-Traps?)

See other training resources at



s/)

Related information

[Model the data for searching \[See page 404\]](#)

Add or manage experts

Summary: You can designate experts for each data source, so users won't get stuck if they need help.

If your users occasionally struggle to answer a specific question, build a chart they have in mind, or find something in their data, you can enable **Ask an Expert**. This lets them request help from someone in your organization.

The request goes to the person in your company who knows the data source very well, and how to search it in ThoughtSpot. The user can see the expert's answer in the **Ask an Expert** interface in ThoughtSpot.

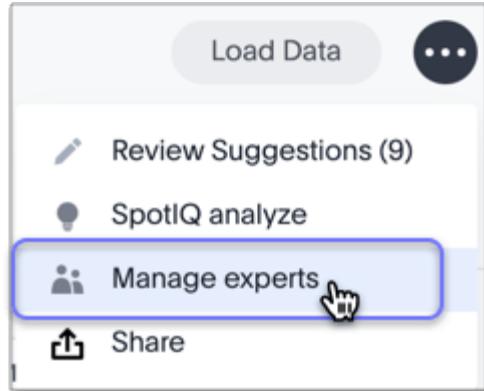
You can see how an end user experiences **Ask an Expert** [here](#) [See page 0].

Add an expert to a data source

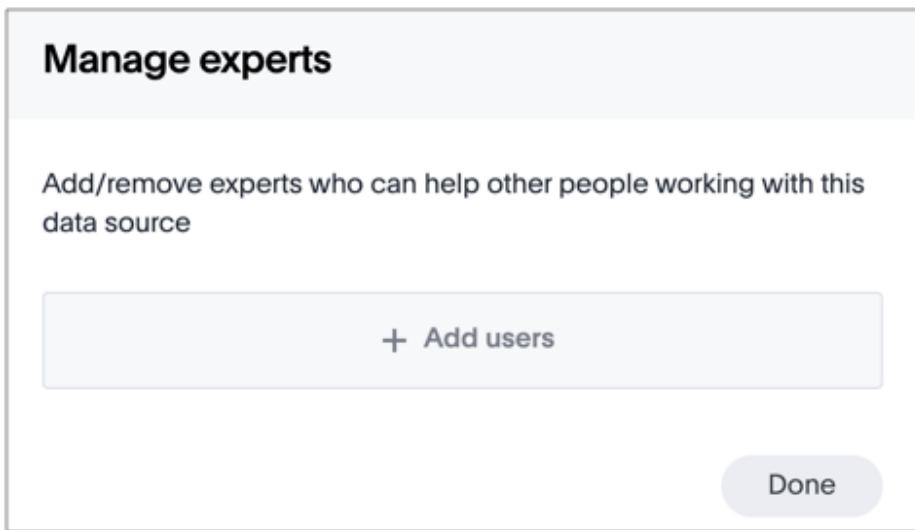
Adding an expert for a data source enables **Ask an Expert** in your ThoughtSpot instance. Until you add an expert, users won't see the prompt to **Ask an Expert** when using that data source.

To add an expert to a data source:

1. Let the expert(s) know to expect user questions to start coming their way in email. If your expert(s) don't have an email in ThoughtSpot, they must [check in ThoughtSpot for requested answers](#) [See page 0].
2. Click the name of the worksheet or table from the **Data** page.
3. Click the three dot icon in the upper right side of the page and select **Manage experts**.



4. Make your selections and click **Done**.



Related information

- Model the data for searching [See page 404]
- Ask an Expert [See page 0]

Link tables using relationships

Summary: Learn how to link tables using relationships.

You can link tables by creating relationships between their columns. Linked tables can be searched together or combined into a worksheet for easy searching. Tables that have no relationship between their columns cannot be combined in a single search.

There are two ways to create relationships between tables:

1. [Create a constraint using TQL \[See page 191\]](#)
2. [Create a relationship through the web interface. \[See page 450\]](#)

The two methods create the same kind of relationship. When creating a relationship between two tables, the columns that form the link must have the same data type. For example, you can build a relationship between two columns that use the `INT` data type. You can also generate a script in TQL that contains all relationships, both the ones created in the web interface, and the ones defined through TQL.

These relationships can be managed either in TQL, or by going to the **Relationships** page when viewing data in the **Data Modeling** section in the ThoughtSpot application. You can view, modify, or delete relationships in either place.

You may create relationships using a mixture of TQL and the web interface, but the relationships you create cannot form a circular relationship, or “cycle”. If you attempt to create a relationship that would complete a cycle, you will see a message stating that the relationship could not be added because it conflicts with another existing relationship.

Join a table or view to another data source

Summary: Learn how to define joins between a table or view and another table, view, or worksheet

Joining a table or view to another table, view, or worksheet creates a relationship that allows them to be searched together. Choose a column to join on that both tables contain (e.g. employee ID or product key). This process creates a [generic join \[See page 191\]](#) between the table or view and the other table, view, or worksheet on the column you specify.

If you want to create a primary key/foreign key relationship, you need to use [TQL \[See page 191\]](#) rather than the web interface.

Note: Defining a generic relationship in the UI rather than using a primary key/ foreign key join through TQL has no impact on performance. However, when creating relationships in the UI, you must ensure that you create it in the right direction: many to one. To create many-to-many joins, or to create joins using >, <, >=, or <=, use TQL.

You must have either the [Can administer ThoughtSpot](#) or the [Can manage data](#) privilege [See page 265] to create a join relationship. If you're not an administrator, you also need edit permissions on the table, view, or worksheet.

See this matrix for information about which joins you can create, and what permissions these joins require.

	Worksheets	View	Materialized View	Imported table (UI)	Table uploaded from backend (tsload) or through DataFlow	Table uploaded through Embrace	View on top of table uploaded through Embrace

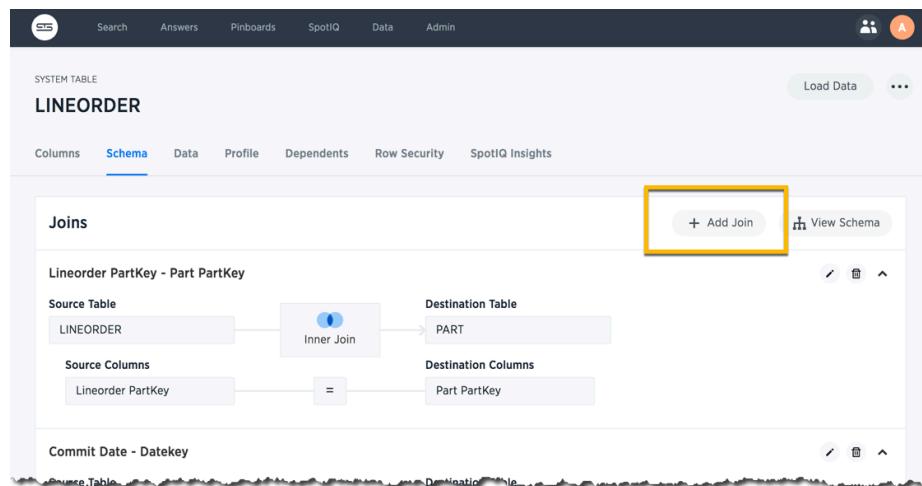
	Necessary permissions:	None	None	None	Can manage data permission to load the table	Admin privileges to access tsload	None	None
Worksheets	Can edit permission on the source Worksheet	x	x	x	✓	x	x	x
View	Can edit permission on the source View	x	✓	✓	✓	✓	x	x
Materialized View	Can edit permission on the source Materialized View	x	✓	✓	✓	✓	x	x
						Note: It is a best practice to create this join through the UI, rather than using TQL .		
Imported table (UI)	Can edit permission on the source table	✓	✓	✓	✓	✓	x	x
Table uploaded from backend (tsload) or through DataFlow	Can edit permission on the source table	x	✓	✓	✓	✓	x	x
					Note: It is a best practice to create this join through the UI, rather than using TQL .			
Table uploaded through Embrace	Can edit permission on the source table, and can manage data permission	x	x	x	x	x	✓	✓
					Note: The two tables must be from the same connection.		Note: The View and the table must be from the same connection.	

View on top of table uploaded through Embrace	Can edit permission on the source View	X	X	X	X	X	✓	Note: The View and the table must be from the same connection.	✓	Note: The two Views must be from the same connection.
--	---	---	---	---	---	---	---	---	---	--

When creating a join between the columns in two data sources, the linked columns must have the same data type, with the same meaning. That is, they must represent the same data. Normally, you can make this kind of link from a fact table column to a column in a dimension table that uniquely identifies a logical entity in your data such as Employee ID for a person, Product ID for a product, or Date Key for a specific date in a date lookup table.

To create a relationship through the Web interface:

1. To find your table or view, click **Data** in the top menu, and choose **Tables or Views**.
2. Click the name of your table or view.
3. Click **Schema**. You will see the list showing existing joins.
4. Click the **+ Add Join** button on the upper right side of the screen.



5. Use the **Map source to destination** dialog to choose the destination table, view, or worksheet you want to join to.

Add Join

Map Source to Destination

Use the form below to map your selected table to a destination table and select specific columns to join.

Source Table LINEORDER	→	Destination Table Select Table
Source Columns Select Column	=	Search table name Users SUPPLIER PRODUCTS Tax Revenue
+ Add Column		

6. Choose the columns you want to join on from the table or view (source) and destination table, view, or worksheet. Click **Next**.

Add Join

Map Source to Destination

Use the form below to map your selected table to a destination table and select specific columns to join.

Source Table LINEORDER	→	Destination Table PRODUCTS
Source Columns Lineorder PartKey	=	Destination Columns Select Column
Search Columns Product Id		

7. Give your join a name and description and click **ADD JOIN**.

8. Repeat these steps until all the joins you want to make have been created.

After creating the join, you may change its name and description by clicking the edit icon. If you want to change the data source or column being joined, you must delete the join and create a new one.

Related Information

- [Constraints \[See page 191\]](#)

Delete a relationship

Summary: You can delete a relationship between tables through the ThoughtSpot application or TQL.

You must have either the **Can administer ThoughtSpot** or the **Can manage data** privilege [See page 265] to delete a relationship. If you're not an administrator, you also need edit permissions on the table, view, or worksheet.

If you created a relationship (join or link) between tables using the Web interface, you can also delete it from the Web interface. But if the relationship was created using TQL, you must also use TQL to delete it.

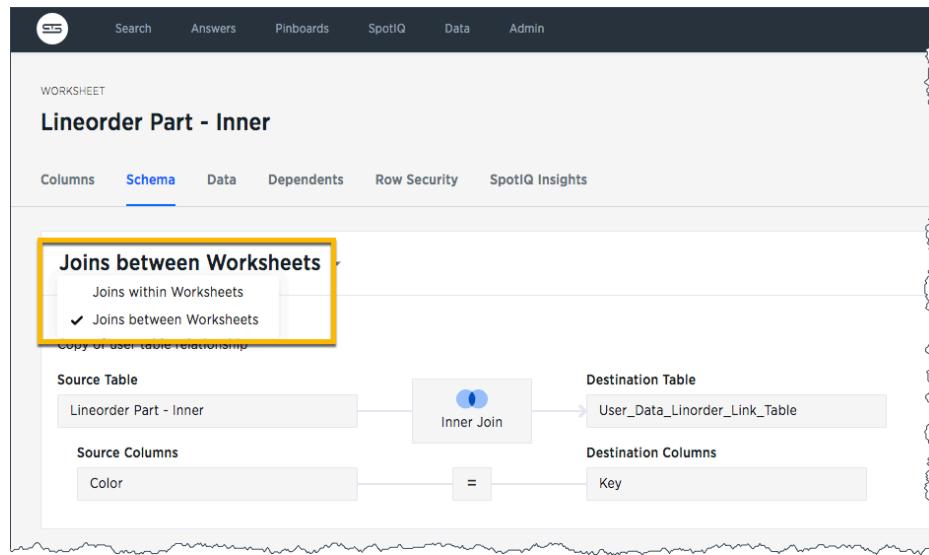
To delete a relationship using TQL [See page 228], use an `ALTER TABLE...DROP CONSTRAINT...` or `ALTER TABLE...DROP RELATIONSHIP...` statement.

To delete a relationship from the Web interface:

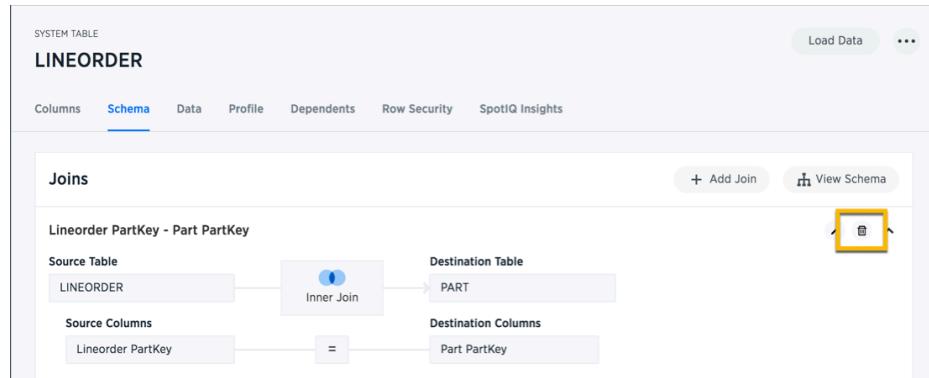
1. Click **Data** on the top navigation bar.
2. Click the name of the data source you from which you want to remove the relationship.
3. Click **Joins**.

The list of existing joins appears.

In a worksheet, you must click **Joins within worksheets** and choose **Joins between worksheets**.



4. Find the relationship you want to delete, and click the delete icon.



5. Repeat these steps until all the joins you want to remove have been deleted.

Related Information

- [Constraints \[See page 191\]](#)

About stickers

Summary: Stickers enable you to create categories for classification of objects, including pinboards, answers, data sources, and worksheets.

You can create stickers to make it easier for people to find data sources and pinboards. Stickers are global in scope. This means that everyone can see the stickers and use them to tag objects. They can also filter lists of objects by sticker. Stickers are often used to designate subject areas, such as sales, HR, and finance, but you can use them any way you like.

Keep in mind these permissions when working with stickers:

- Only administrators can create stickers.
- Anyone who has **edit** access to an object can apply a sticker to it.
- Anyone can filter by a sticker.

Create a sticker

Only administrator users can create stickers. Anyone can apply the stickers you create, or use them as filters when selecting from a list of sources or pinboards.

To create a sticker:

1. Navigate to the **Manage Data** or **Pinboards** screen using the icons in the top navigation bar.
2. Choose the currently selected sticker, scroll to the bottom of the list, and click **+ Add**.

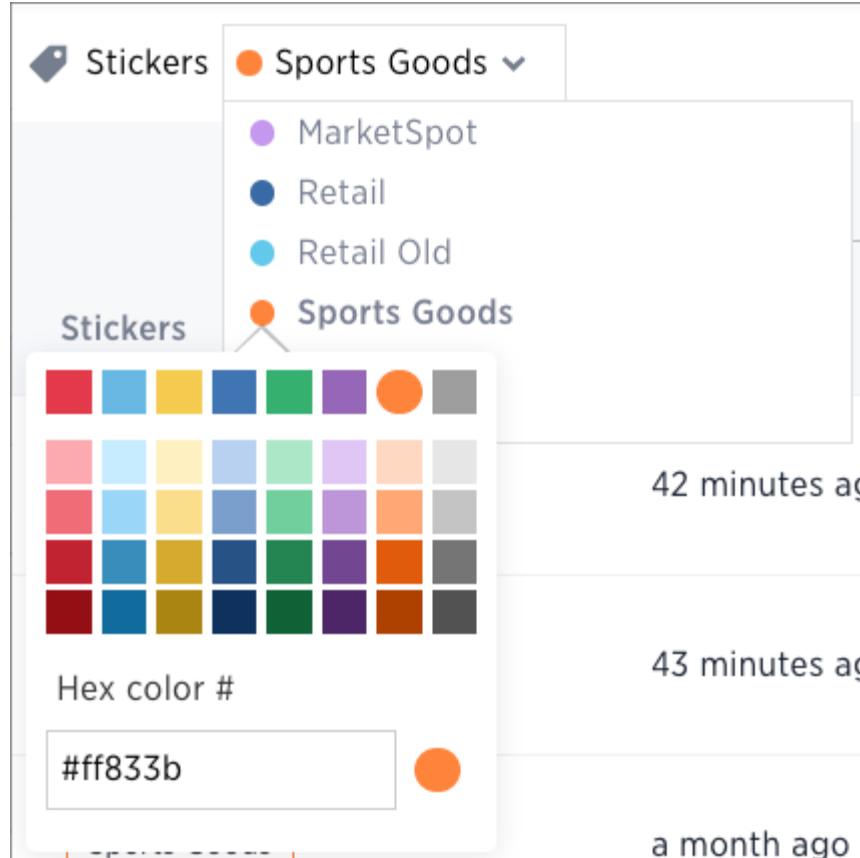
The screenshot shows the ThoughtSpot interface with the 'Stickers' tab selected. A dropdown menu is open, showing several categories: MarketSpot (purple), Retail (dark blue), Retail Old (light blue), and Sports Goods (orange). The 'Sports Goods' category is currently selected. Below the dropdown, there is a '+ Add' button with a red border. A table lists four items, each with a 'Sports Goods' sticker and a timestamp:

Name	Stickers	Timestamp
ThoughtSPORT Overview	Sports Goods	9 minutes ago
Advanced Analysis with R	Sports Goods	9 minutes ago
ThoughtSPORT Analysis Overview of Advanced Formulas in Thoughtspot	Sports Goods	a month ago
Comparative Analysis	Sports Goods	a month ago

3. Type the name for the new sticker.
4. You can change the name of a sticker by clicking the edit icon next to its name.

The screenshot shows the ThoughtSpot interface with the 'Stickers' tab selected. A dropdown menu is open, showing options: 'Edit name' (highlighted in grey) and 'Remove sticker'. Below the dropdown, there is a list of categories with their corresponding color circles: MarketSpot (purple), Retail (dark blue), Retail Old (light blue), and Sports Goods (orange). A small edit icon is also visible.

5. You can change the color of a sticker by clicking the color circle next to its name.

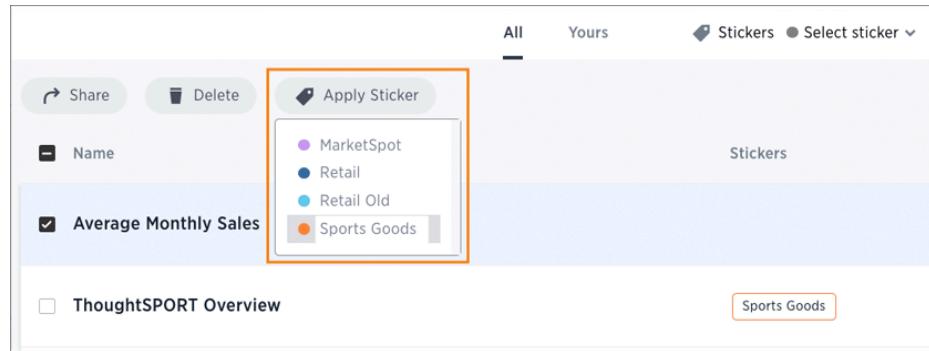


Apply a sticker

Only administrators create stickers, but anyone with edit privileges can tag an object with a sticker.

To tag an object with a sticker:

1. From the top menu, choose Answers, Pinboards, or Data.
2. Find the item(s) you want to tag in the list, and check the box next to its name.
3. Click the apply sticker icon and choose one from the list. You can apply as many stickers as you like to an object.



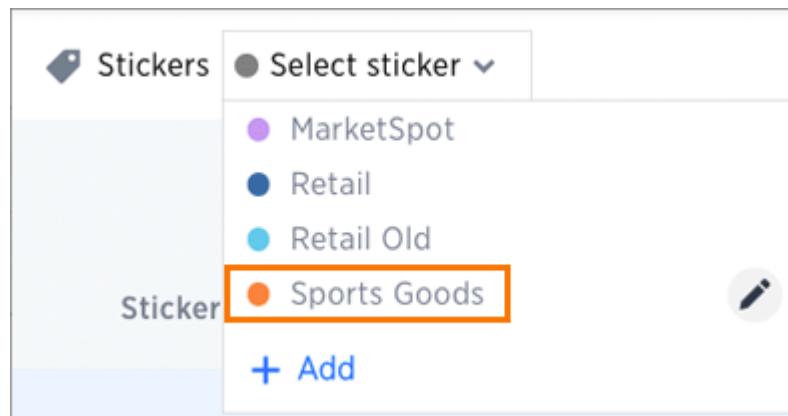
Filter by stickers

Whenever you are selecting objects from a list, you can filter by sticker to find what you're looking for.

Anyone can use stickers to filter lists of pinboards or data sources. You can also filter by sticker when selecting data sources.

To filter by sticker:

1. From the top menu, choose **Answers**, **Pinboards**, or **Data**.
2. Click **Select sticker**, and select the name of the sticker you want to filter by.



Create and use worksheets

Summary: Worksheets are flat tables created by joining columns from a set of one or more tables or imported datasets.

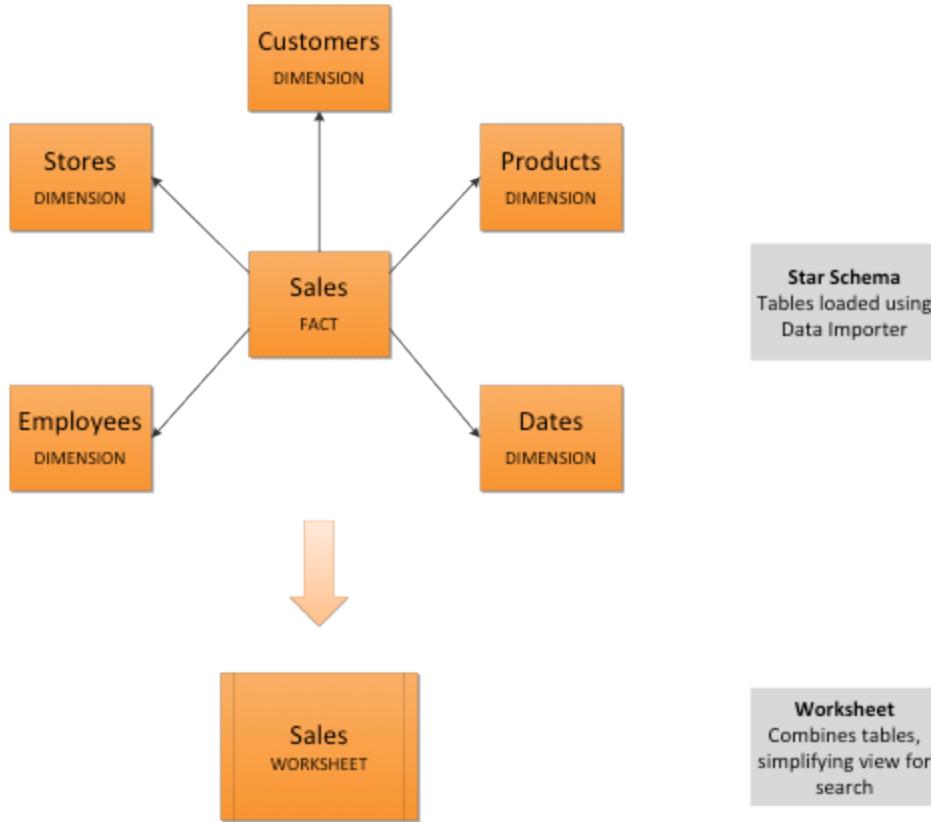
After [modeling your data \[See page 404\]](#), create worksheets to make searching easier. For example, a sales executive might need to search for information about retail sales. This data might be contained in several tables (sales, customers, products, stores, etc.), with foreign key relationships between them. An administrator who is familiar with the data model can create a retail sales worksheet, that combines all of the related fact and dimension tables into a single, easy-to-use view, and share it with the sales executive. This provides access to the data without requiring an understanding of how it is structured.

Guidelines for worksheets

Users are often unfamiliar with tables and how they are related to one another. A worksheet groups multiple related tables together in a logical way. You might use a worksheet for these reasons:

- To pre-join multiple tables together.
- To give a user or group access to only part of the underlying data.
- To include a derived column using a formula.
- To rename columns to make the data easier to search.
- To build in a specific filter or aggregation.
- To give users a filtered set of data to search.

Typically, you create one worksheet for each set of fact and dimension tables. For example, you may have a sales fact table and an inventory fact table. Each of these fact tables shares common dimensions like date, region, and store. In this scenario, you would create two worksheets: sales and inventory. The following diagram depicts the workflow for creating the sales worksheet.



The process for creating a worksheet is:

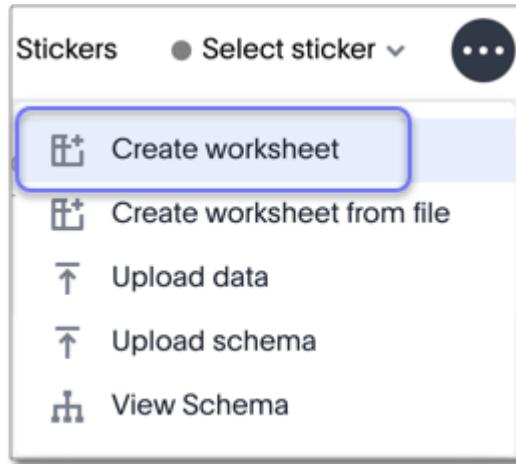
1. Decide which tables to use for the worksheet.
2. [Create a new Worksheet \[See page 463\]](#). If the worksheet already exists in another cluster, you can migrate it [using a flat TSL file \[See page 521\]](#).
3. Add sources (tables) to the worksheet.
4. Choose the [worksheet join rule \[See page 477\]](#).
5. Select the columns to include.
6. Optionally [modify the join types \[See page 488\]](#) within the worksheet.
7. Optionally [create formulas \[See page 469\]](#).
8. Optionally [create worksheet filters \[See page 472\]](#).

9. Save the worksheet.
10. Share the worksheet with groups or users [See page 318].

Create a worksheet

To create a new worksheet:

1. Click **Data**, on the top navigation bar.
2. Click the ellipsis icon  , and select **Create worksheet**.



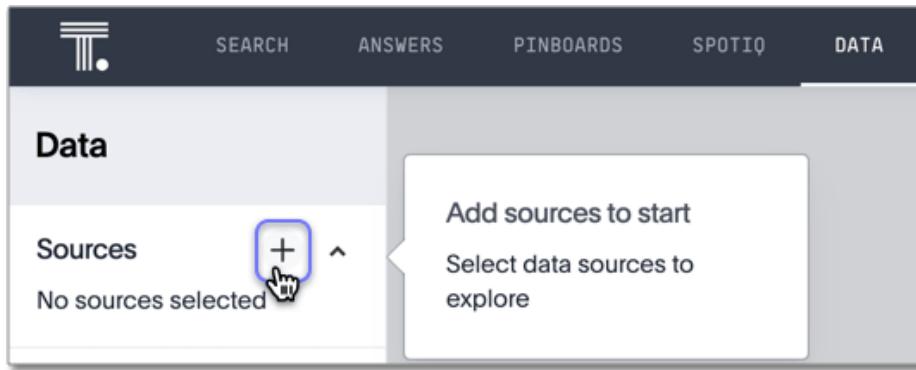
3. Add sources and columns [See page 463].

Add sources and columns to a worksheet

After creating a worksheet, you need to add the sources that contain the data. A source is another name for a table. The sources you choose are typically related to one another by foreign keys.

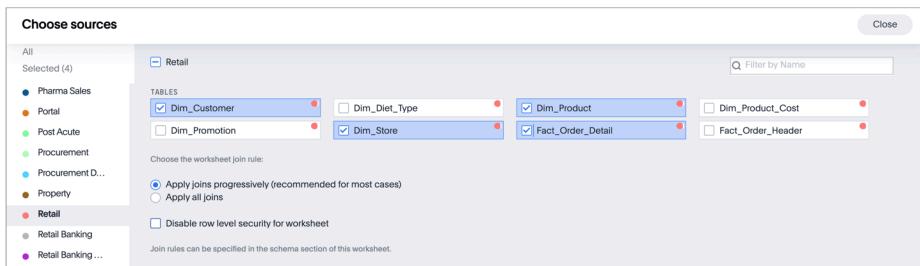
To add sources to your worksheet, follow these steps. The worksheet creation UI also guides you through the process.

1. Click the + icon next to **Sources**.



2. Check the box next to each of the sources you want to include in the worksheet. You can search for specific Views, imported data, or tables. You can also select every data source that has a specific sticker, like **Retail**.

Note that the list of sources only shows the data sources on which you have view or edit privileges.

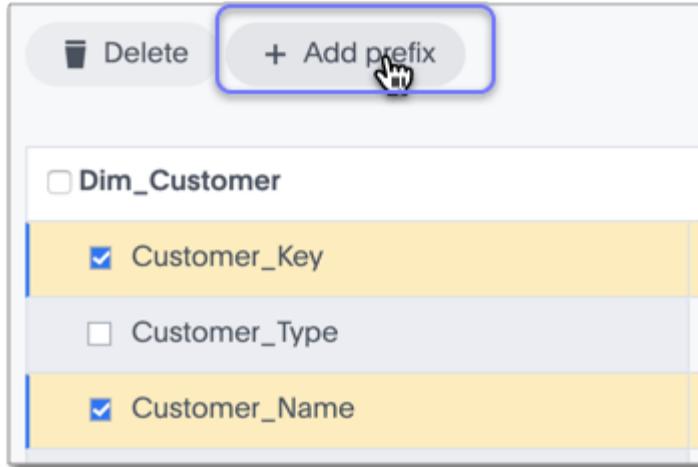


3. Choose the [worksheet join rule \[See page 477\]](#). Either **apply joins progressively** or **apply all joins**. Applying joins progressively speeds up performance.
4. If you want to disable [Row Level Security \[See page 351\]](#) for this worksheet, check the checkbox to disable it.
5. Click **CLOSE** to save your changes.
6. Expand the table names under **Sources** and select the columns to add to the worksheet, by doing any of the following:
 - a. To add all of the columns from a table, click the table name and click **+ Add Columns**.

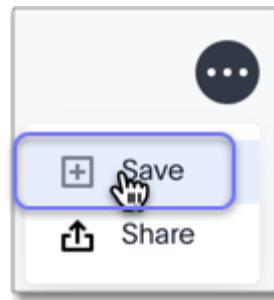
- b. To add a single column, double-click its name.
- c. To add multiple columns, Ctl+click each column you want to add and click **+** **Add Columns**.

Note that after you add a column, non-related tables (those without a primary/foreign key relationship) become hidden. If you are working with two tables that should be related, but are not, you can [add a relationship between them \[See page 449\]](#).

7. (Optional) [Modify the join types \[See page 488\]](#) within the worksheet.
8. (Optional) [Create formulas \[See page 469\]](#).
9. (Optional) [Create worksheet filters \[See page 472\]](#).
10. Click the ellipsis icon  , and select **Save**.
11. In the Save Worksheet window, enter a name and description for your worksheet and click **SAVE**.
12. (Optional) Click each column name and enter a more user-friendly name for searching. You can tab through the list of columns to rename them quickly.
13. (Optional) If you want to add a prefix to the name of several columns, select them, click the **Add prefix** button, and type in the prefix.



14. Click the ellipsis icon ..., and select **Save**.



15. Share your worksheet [See page 318], if you want other people to be able to use it.

Where to go next

- [How the worksheet join rule works \[See page 477\]](#)

Use the worksheet join rule to specify when to apply joins when a search is done on a worksheet. You can either apply joins progressively, as each search term is added (recommended), or apply all joins to every search.

Edit or rename Worksheet

Summary: Anyone with the proper permissions can edit a Worksheet.

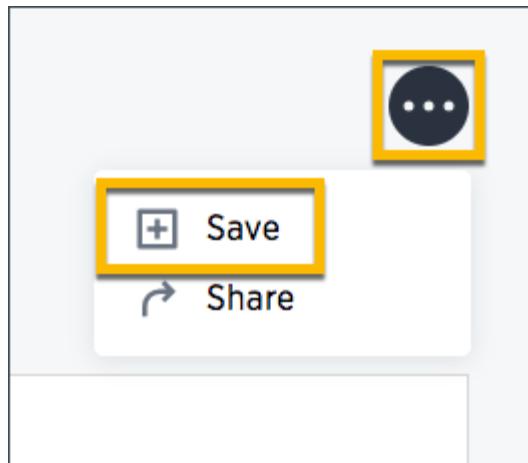
If you created a Worksheet, or you have edit permissions on it, you can make changes such as adding sources and columns, adding or editing formulas, changing relationships, and changing column names. You can also rename a Worksheet or change its description.

Alternatively, you can export a Worksheet as a flat file in [TSL \[See page 535\]](#), ThoughtSpot's Scripting Language, edit the file, and import the file to the same or another cluster. See [Migrate or restore objects \[See page 521\]](#).

Edit a Worksheet

To edit a Worksheet:

1. Click **Data** on the top navigation bar.
2. Click the name of the worksheet you want to edit.
3. Click the **Edit** button in the upper-right side of the screen.
4. Make your changes to the worksheet.
5. Click the the ellipsis icon  , and select **Save**.



Rename a worksheet or table

You can change a worksheet or table name from the ThoughtSpot application.

To change the name of a worksheet or table:

1. Click **Data**, on the top navigation bar.
2. Find the worksheet or table you want to rename and click its name.
3. Click the current name, and enter a new name.

Related information

- [Change the join rule for a worksheet \[See page 481\]](#)
- [Add joins between a worksheet and other data \[See page 483\]](#)
- [Modify table joins within a worksheet \[See page 488\]](#)

Create a formula in a worksheet

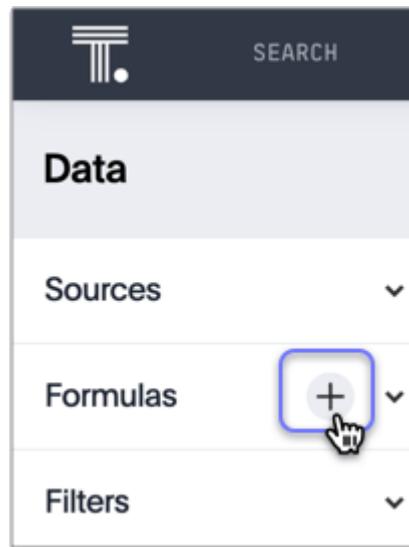
Summary: You can define formulas and use them to create derived columns in worksheets.

You create formulas by combining standard functions and operators, column names, and constant values.

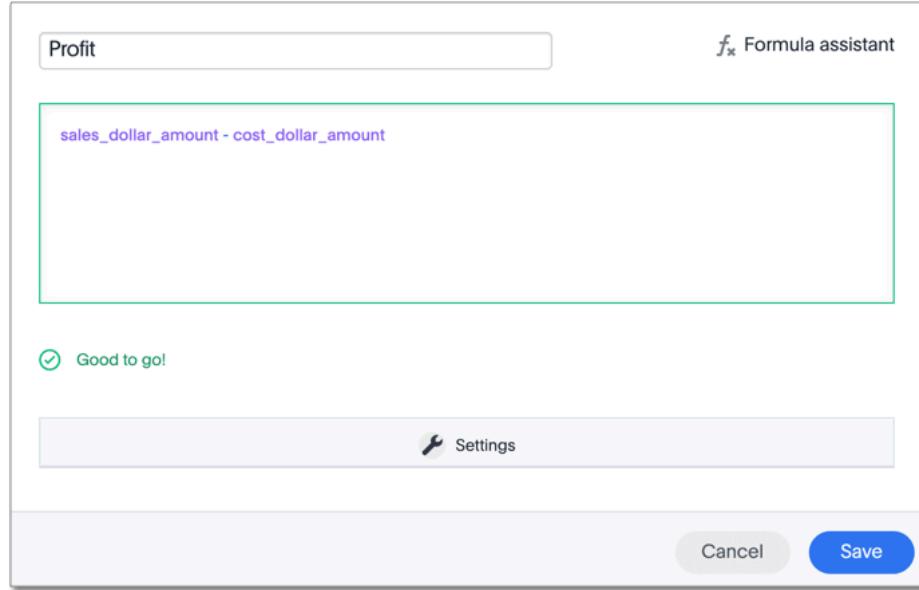
Anyone who can create a worksheet can add a formula to it. Formulas are not reusable; the formula you create is associated only with the worksheet it belongs to. A complete list of available formulas and examples of each is available in the [Formula function reference \[See page 0\]](#).

You can create a formula in a worksheet by using the Formula Builder. When you do this, the result of the formula gets added to the worksheet as a column. Use these steps to create a formula:

1. [Create a new worksheet \[See page 461\]](#), or [edit an existing one \[See page 467\]](#).
2. Click the + button next to **Formulas**.

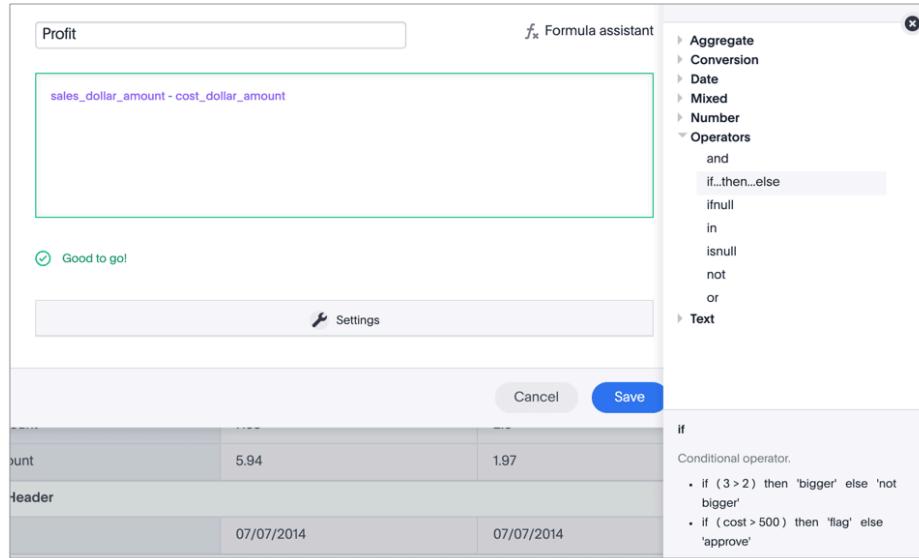


3. Type your formula in the Formula Builder.



Note: Formula elements are color coded by type and can include the formula operators and functions (blue), the names of columns (purple), and/or constants (black).

4. You can see a list of formula operators with examples by clicking on **Formula Assistant**.

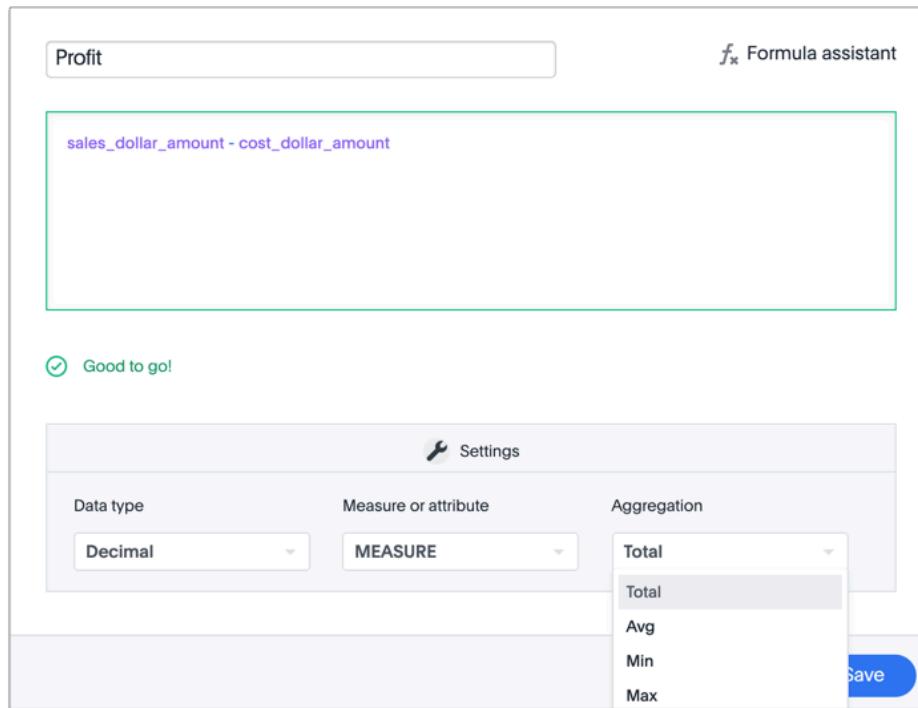


5. If you want to change what your formula returns, use the **Advanced settings**.

Depending on your formula, you may be able to change:

- Data type

- ATTRIBUTE or MEASURE
- Aggregation type



6. Name the formula by clicking on its title and typing the new name. Click **Save**.

Use aggregate formulas as Worksheet filters

Starting in ThoughtSpot release 6.2, you can use an aggregate formula as a Worksheet filter. This is useful when, for example, you only want your results to show a measure when the related attribute is greater than some number, or vice versa. You may only want to see `sales` when the `unique customer count` is greater than 1, or you may only want to see a `customer` if the associated `sales` is greater than 0. Rather than add that formula to every search, you can create a filter at the Worksheet level.

To add an aggregated formula to a Worksheet, follow these steps:

1. Create an [aggregate formula \[See page 0\]](#) in a Worksheet, as shown in steps 1-6 above.
2. [Add the formula to Worksheet as a filter \[See page 472\]](#), and specify conditions in which it should apply.

Add a filter to a worksheet

Summary: You can add filters to a worksheet to limit the data users can access from the worksheet.

You can add filters to a worksheet to limit the data it contains. This is useful when you have underlying tables that store more data than is necessary for the types of analyses the worksheet is intended for.

You can also use worksheet filters to provide data security, when you want different groups of users to be able to see different data without relying on row level security.

Note that starting in release 6.2, you can export and import Worksheets with filters [using a flat TSL file \[See page 521\]](#), which allows for optional modification of any Worksheet properties, including filters. If you are on a release prior to 6.2, you can migrate Worksheets using a flat file format, but you cannot migrate Worksheets with filters.

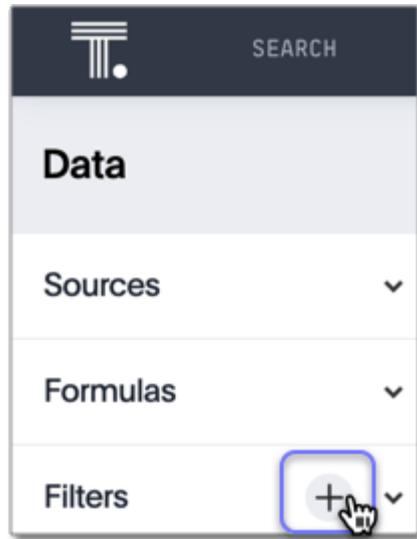
Refer to [Migrate and restore objects \[See page 521\]](#) to add a filter to a Worksheet through the TSL file.

To add a filter to a worksheet through the user interface:

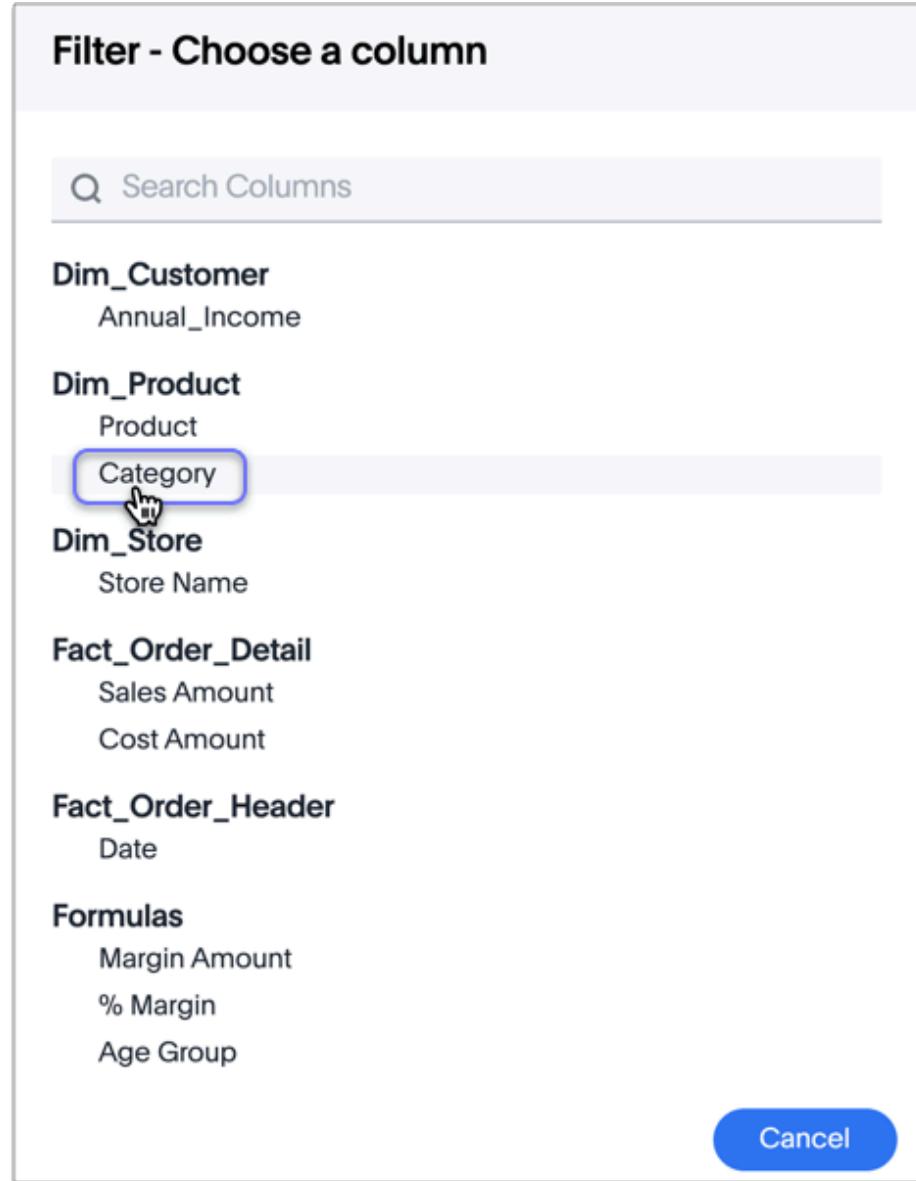
1. Click **Data** in the top menu bar, find your worksheet, and click its name.
2. Click the **Edit Worksheet** button.

COLUMN NAME	DESCRIPTION	DATA TYPE	COLUMN TYPE	ADDITIVE	AGGREGATION	HIDDEN	SYNONYMS	SPOTIQ PREFERENCE
Sales Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Sales Amt, Reve...	DEFAULT
Cost Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Click to edit	DEFAULT
Store Name	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT
Product	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT

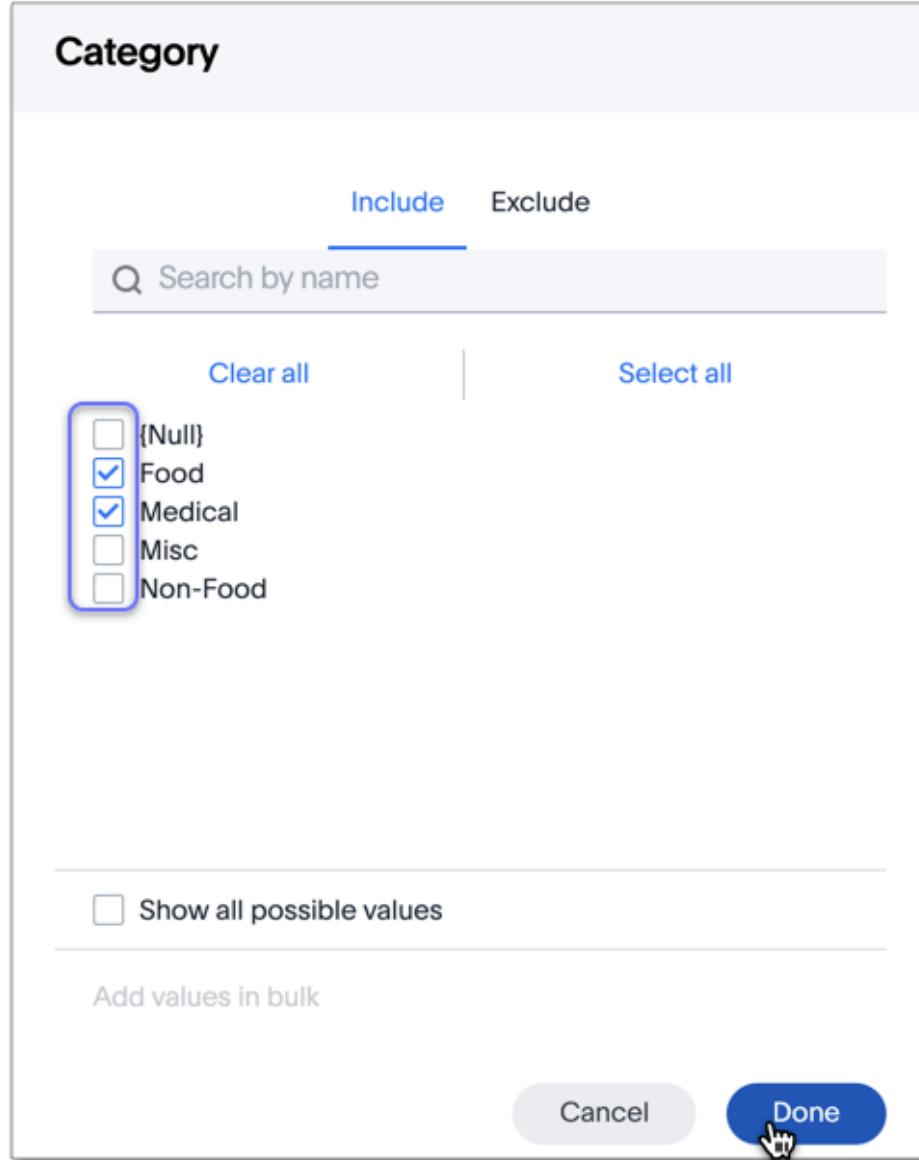
3. Hover over **Filters** on the left menu and click **+**.



4. Choose the column you want to filter on.



5. Select the values to include in your answer.



6. If you want to exclude values, click **Exclude** and choose values to exclude.

7. Click **Done**.

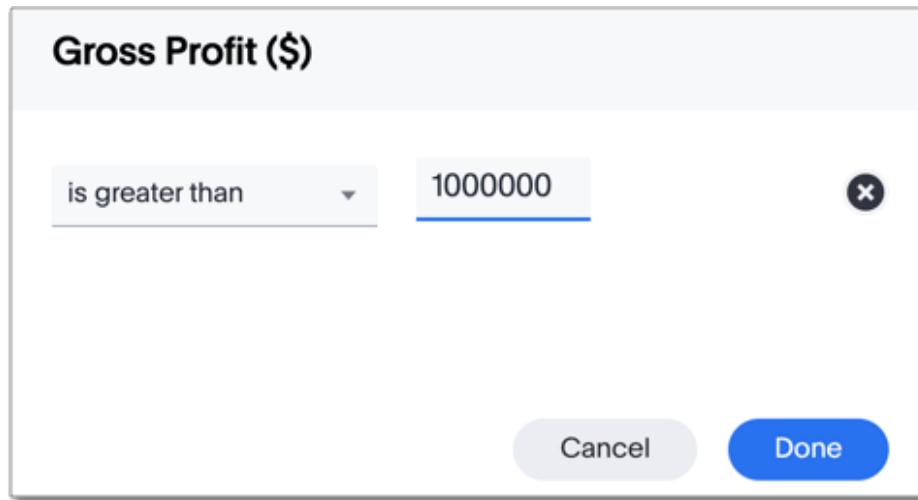
If there are too many values, you can use the filter search bar to find the ones you want.

Use aggregate formulas as Worksheet filters

Starting in ThoughtSpot release 6.2, you can use an [aggregate formula \[See page 0\]](#) as a Worksheet filter. This is useful when, for example, you only want your results to show a measure when the related attribute is greater than some number, or vice versa. You may only want to see `sales` when the `unique customer count` is greater than 1, or you may only want to see a `customer` if the associated `sales` is greater than 0. Rather than add that formula to every search, you can create a filter at the Worksheet level.

To add an aggregated formula to a Worksheet, follow these steps:

1. [Create an aggregate formula in a Worksheet \[See page 469\].](#)
2. Add the formula to Worksheet as a filter, as shown in steps 1-7 above. Instead of choosing a column to filter on, search for and choose your aggregate formula.
3. Specify conditions in which the filter should apply. Here, we have chosen to only display `gross profit` when it is greater than \$1,000,000. The following conditional statements are acceptable: `is greater than`, `is greater than or equal to`, `is less than`, `is less than or equal to`, `is between`, `is in`, or `is not in`.



4. Click **Done**, and **save** the Worksheet.

How the worksheet join rule works

Summary: The worksheet join rule specifies when to apply joins when searching on a worksheet.

Use the worksheet join rule to specify when to apply joins when a search is done on a worksheet. You can either apply joins progressively, as each search term is added (recommended), or apply all joins to every search.

Understand progressive joins

Often, a worksheet includes several dimension tables and a fact table. With progressive joins, if your search only includes terms from the fact table, you can see all of the rows that satisfy your search. But as you add terms from dimension tables, the total number of rows shown may be reduced, as the joins to each dimension table are applied. It works like this:

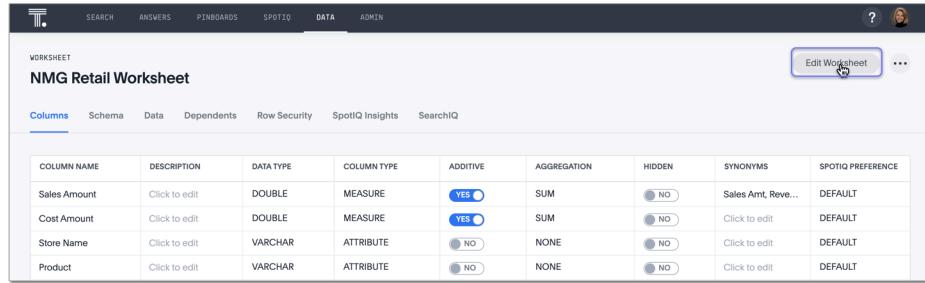
- If you choose **Apply joins progressively (recommended for most cases)**, joins are only applied for tables whose columns are included in the search.
- If you choose **Apply all joins**, all possible joins are applied, regardless of which tables are included in the search.

When using **Apply joins progressively**, the number of rows in a search using the worksheet depends on which tables are part of the search. The worksheet acts like a materialized view. This means that it contains the results of a defined query in the form of a table. So if a particular dimension table is left out of the search, its joins are not applied.

Apply progressive joins

To specify how the Worksheet should apply joins, follow these steps:

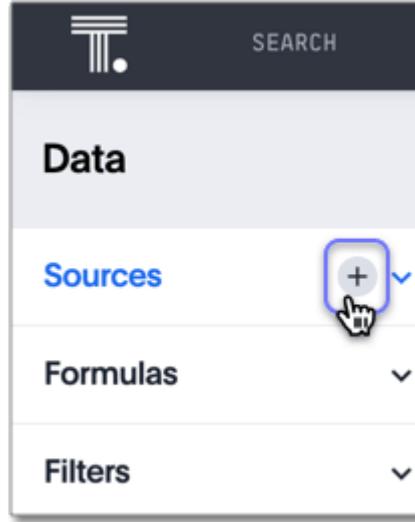
1. Navigate to the Worksheet you would like to edit.
2. Select **Edit Worksheet** in the top right corner.



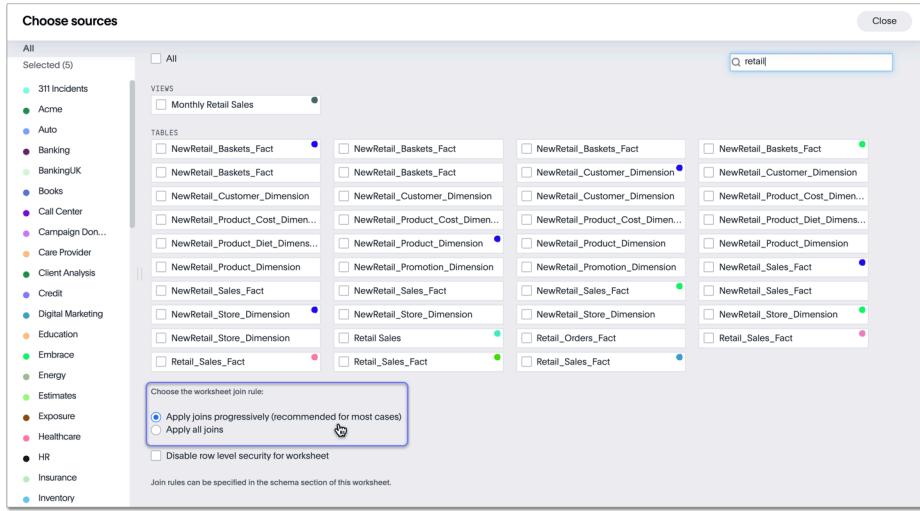
COLUMN NAME	DESCRIPTION	DATA TYPE	COLUMN TYPE	ADDITIVE	AGGREGATION	HIDDEN	SYNONYMS	SPOTIQ PREFERENCE
Sales Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Sales Amt, Reve...	DEFAULT
Cost Amount	Click to edit	DOUBLE	MEASURE	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Click to edit	DEFAULT
Store Name	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT
Product	Click to edit	VARCHAR	ATTRIBUTE	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT

If you do not see this option, you may not have editing privileges on this Worksheet. Request edit access from your administrator, or the owner of the Worksheet.

3. Click the + button to add sources and specify join application, or to specify how to apply joins on the Worksheet's existing sources.



4. Select the sources you want to specify join application for.
5. Scroll down to the bottom of the **Choose sources** modal, and under **Choose the worksheet join rule**, select either **apply joins progressively** or **apply all joins**.



Rule-Based Row Level Security (RLS) with worksheets

With Rule-Based RLS, you need to protect every table that contains any sensitive data. To do this, you can grant access by creating explicit row level security rules on each of the underlying tables which contain data that row level security should apply to.

When creating the row level security rules for a table that's part of a worksheet, you aren't limited to referencing only the columns in that table. You can specify columns from other tables in the worksheet as well, as long as the tables are joined to the table you're creating the rule on. Then, when creating a worksheet on top of them, the behavior is consistent regardless of the worksheet join rule you choose. Users will never be able to see data they should not, regardless of what their search contains.

Imagine you have a worksheet that contains a `Sales` fact table, and `Customer` and `Product` dimensions that are joined on `Customer SSN` and `Product Code` columns. In order to secure the `Sales` table, you can use `Customer Name` from the `Customer` column to create a row level security rule.

How joins are applied with chasm traps

When working with worksheets and row level security, you need to understand how joins are applied.

This is particularly important with chasm trap schemas. For chasm trap schemas, if row level security is only set on one of the tables, people could see data they should not see if the scope of their search does not include that table. (this protects them from having people see the wrong things if they have chasm trap).

For chasm trap *worksheets*, progressive and non-progressive joins do not apply. There is an entirely different methodology for how worksheet joins on a chasm trap schema work with row level security. So you can safely ignore that setting.

Change the join rule or RLS setting for a worksheet

Summary: As long as you have permissions to edit a worksheet, you can always go into it and set a different join or RLS rule.

If you find that the charts and tables built on a worksheet contain a large number of null values (which display as {blank} in the web browser), you can fix this by changing the [internal joins \[See page 488\]](#) for the worksheet.

If you have the **Can administer ThoughtSpot** privilege or the **Can manage data** plus edit privilege on a worksheet, you can edit the worksheet and change its RLS or other key settings.

Change/configure a worksheet

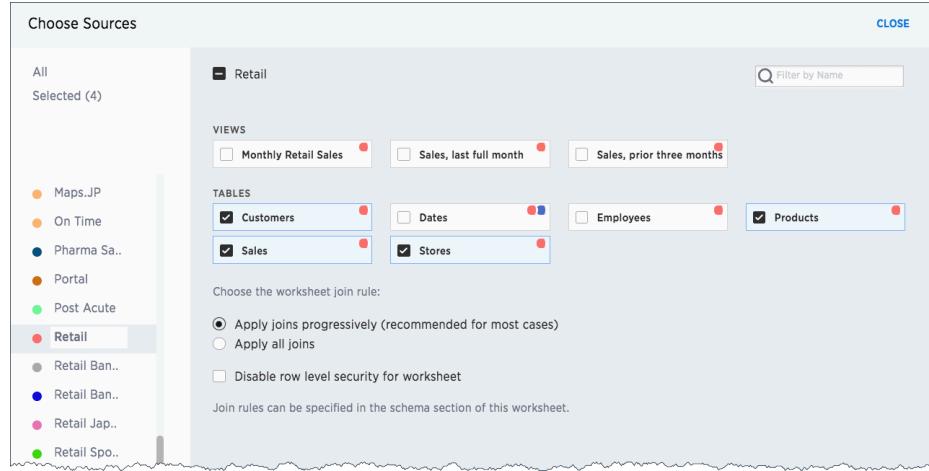
Before working through this procedure, make sure you are familiar with how the following affect data:

- [internal worksheet joins \[See page 488\]](#)
- [worksheet join rule \[See page 477\]](#)
- [row level security \(RLS\) \[See page 356\]](#)

To configure these values for a worksheet:

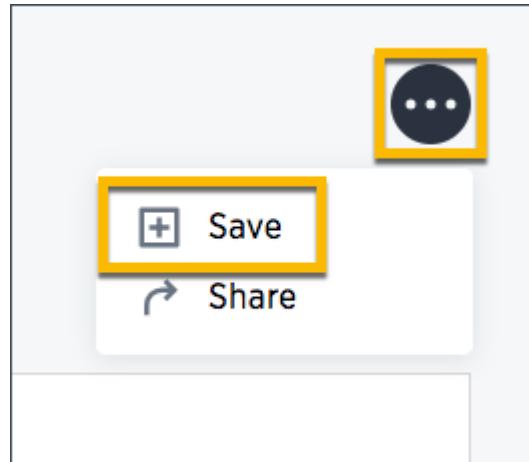
1. Click **Data** on the top navigation bar and then on **Worksheets**.
2. Click the name of the worksheet you want to edit from the list.
3. Click the **Edit Worksheet** button in the upper right hand side of the screen.
4. Click the **+** icon next to **Sources**.
5. Scroll to the bottom of the page.
6. Configure the worksheet join rule and RLS setting as needed.

Change the join rule or RLS setting for a worksheet



7. Click **CLOSE**.

8. Click the ellipses icon , and select **Save**.



Join a worksheet to another data source

Summary: Learn how to define joins between a worksheet and a table or view.

Joining a worksheet to a table or view creates a relationship that allows them to be searched together.

Choose a column to join on that both data sources contain (e.g. employee ID or product key). This process creates a [generic join \[See page 191\]](#) between the worksheet and the table or view on the column you specify.

Note: Defining a generic relationship in the UI rather than using a primary key/ foreign key join through TQL has no impact on performance. However, when creating relationships in the UI, you must ensure that you create it in the right direction: many to one. To create many-to-many joins, or to create joins using >, <, >=, or <=, use TQL.

See this matrix for information about which joins you can create, and what permissions these joins require.

	Worksheets	View	Materialized View	Imported table (UI)	Table uploaded from backend (tsload) or through DataFlow	Table uploaded through Embrace	View on top of table uploaded through Embrace
Necessary permissions:	None	None	None	Can manage data permission to load the table	Admin privileges to access tsload	None	None
Worksheets	Can edit permission on the source Worksheet	X	X	X	✓	X	X

View	Can edit permission on the source View	X	✓	✓	✓	✓	✓	X	X
Materialized View	Can edit permission on the source Materialized View	X	✓	✓	✓	✓	✓	X	X
Imported table (UI)	Can edit permission on the source table	✓	✓	✓	✓	✓	✓	X	X
Table uploaded from backend (tsload) or through DataFlow	Can edit permission on the source table	X	✓	✓	✓	✓	✓	X	X
Table uploaded through Embrace	Can edit permission on the source table, and can manage data permission	X	X	X	X	X	✓	Note: The two tables must be from the same connection.	Note: The View and the table must be from the same connection.
View on top of table uploaded through Embrace	Can edit permission on the source View	X	X	X	X	X	✓	Note: The View and the table must be from the same connection.	Note: The two Views must be from the same connection.

Tip: Defining a generic relationship in the UI rather than using a primary key/ foreign key join through TQL has no impact on performance. However, when creating relationships in the UI, you must ensure that you create it in the right direction: many to one. To create many-to-many joins,

or to create joins using `>`, `<`, `>=`, or `<=`, use TQL.

You must have either the **Can administer ThoughtSpot** privilege or the **Can manage data** privilege to create a join relationship. If you're not an administrator, you also need edit permissions on the table, view, or worksheet.

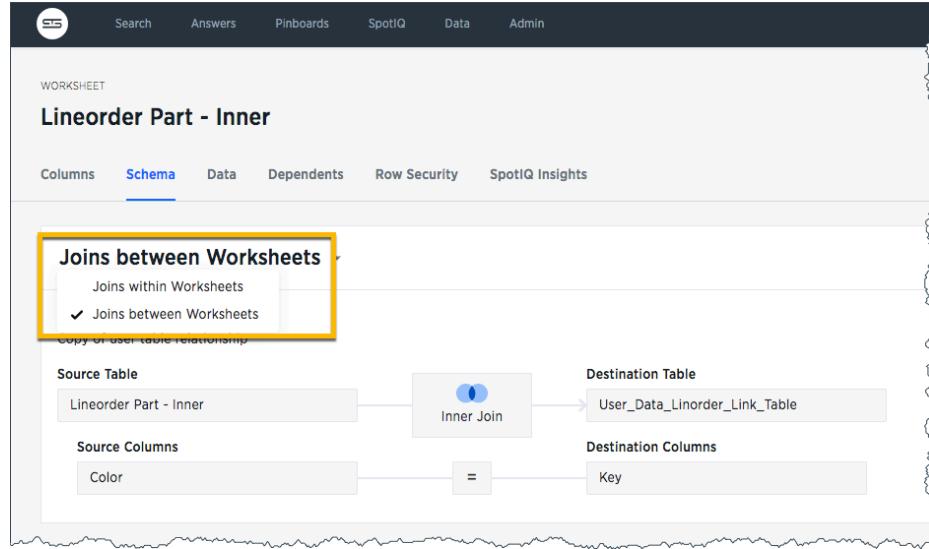
When creating a join between the columns in two data sources, the columns being linked must have the same data type, with the same meaning. That is, they must represent the same data.

To create a relationship through the Web interface:

1. To find your worksheet, click **Data** in the top menu, and choose **Worksheets**.
2. Click the name of your worksheet.
3. Click **Joins**.

The list of existing joins within the worksheet appears.

4. To view the joins between the worksheet and other data sources, click **Joins within worksheets**, and choose **Joins between worksheets**.



5. Click **+ Add Join** on the upper right side of the screen.

WORKSHEET
Lineorder Part - Inner

Columns Schema Data Dependents Row Security SpotIQ Insights

Joins between worksheets **+ Add Join** View Schema

You have not created any joins yet
This table is not joined with any other table.

6. Use the **Map source to destination** dialog to choose the destination table or view for the other end of the join.

Add Join

Map Source to Destination
Use the form below to map your selected table to a destination table and select specific columns to join.

Source Table
Lineorder Part - Inner

Destination Table
User_Data_Linorder_Link_Table

Source Columns
Select Column

+ Add Column

Cancel NEXT

7. Choose the matching columns for the join from the worksheet (source) and destination data source.

Click **Next**.

Add Join

Map Source to Destination

Use the form below to map your selected table to a destination table and select specific columns to join.

Source Table	Destination Table
Lineorder Part - Inner	User_Data_Linorder_Link_Table

Source Columns Destination Columns

Commit Date Select Column

+ Add Column Search Columns

Created Updated

ncel NEXT

8. Give your join a name and description, and click **ADD JOIN**.

9. Repeat these steps to create all joins.

After creating the join, you may change its name and description by clicking the edit icon. If you want to change the data source or column being joined, you must delete the join and create a new one.

Related Information

- Constraints [See page 191]

Modify joins within a worksheet

Summary: Learn how to change the join type between the tables within a worksheet.

When you create a worksheet, you select a [join rule \[See page 477\]](#). The join rule works together with the joins defined within the worksheet determine how the tables that make up the worksheet are joined, and how those joins behave when searching on the worksheet.

Starting with ThoughtSpot version 5.0, you are not limited to just one join rule for the entire worksheet.

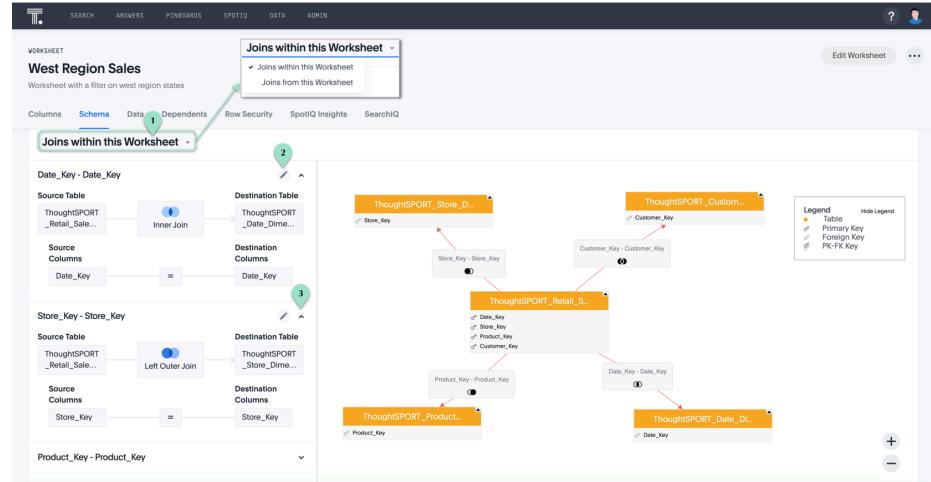
You can define different types of joins for each join between tables in a worksheet. By default, each of these individual table joins uses an inner join. But you can override this at the individual join level.

You must have either the **Can administer ThoughtSpot** privilege or the **Can manage data** privilege to modify joins within worksheets.

Modify a join

To modify the join types within a worksheet, follow these steps:

1. Click the **Data** tab in the top menu.
2. Under **Data Objects**, select ****Worksheets**.
3. Click the name of the Worksheet. Here, we selected the worksheet *West Regional Sales*.
4. At the top of the Worksheet, click **Joins**.
5. The join information and the schema for the worksheet appear.
 - The schema representation includes the join type.
 - You can see the list of *Joins within the worksheet*, which include all joins between the underlying tables, both [created using TQL \[See page 191\]](#) and [created in the browser \[See page 450\]](#).
 - The fact table is always on the left side of the join, and it appears on the left side.



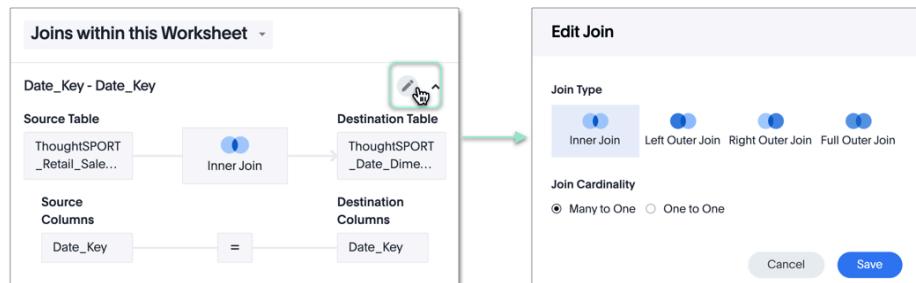
Note the available actions for the join view.

Legend Action

1. See either *Join within this Worksheet (Default)*, or *Joins from this Worksheet*.
2. Click the pencil icon to edit the join.
3. Click the up arrow icon to collapse the join detail.

6. Select the join you plan to modify, and click the **Edit** (pencil) icon next to it.

7. In the **Edit Join** window modal, make the necessary changes:



- Under **Join Type**, select one of *Inner Join*, *Left Outer Join*, *Right Outer Join*, or *Full Outer Join*.
- Under **Join Cardinality**, select either *Many to One*, or *One to One*.

Click **Save**.

Related Information

- [Join rule \[See page 477\]](#)
- [Create joins using TQL \[See page 191\]](#)
- [Create join relationships in the browser \[See page 450\]](#)

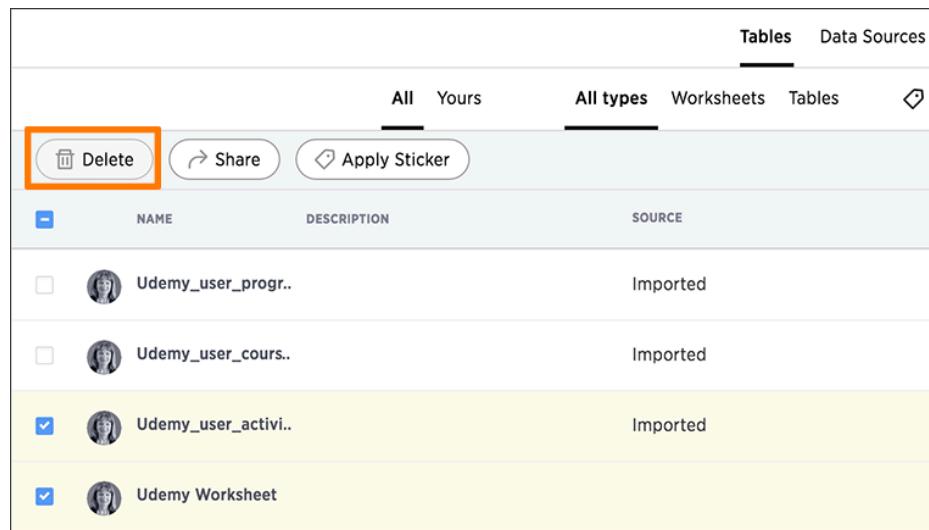
Delete a worksheet or table

Summary: When you try to delete a worksheet or a table, you see a message listing any dependent objects that must be removed first.

ThoughtSpot checks for dependencies whenever you try to remove a table or worksheet. A list of dependent objects appears, and you can click them to delete them or remove the dependency. Then, you can remove the table or worksheet.

To delete a worksheet or table:

1. Click **Data**, on the top navigation bar.
2. Find the worksheet or table you want to remove in the list, and check the box next to its name.
3. Click the **Delete** icon.



If you are attempting to delete a data source with dependent objects, the operation will be blocked. You will see a warning, with a list of dependent objects with links.

4. Click the link for an object to modify or delete it.

When all its dependencies are removed, you will be able to delete the data source.

Cannot delete

The following object(s) depend on "Sporting Goods Retail Worksheet". You must delete them to delete "Sporting Goods Retail Worksheet".

- [Sales by Store - Last 30 Days](#) (Answer)
- [Shopping Trend of Members vs Non-Members](#) (Answer)
- [Sales Breakdown by Mode of Payment](#) (Answer)
- [Moving Sum of All Sales - Last 30 days](#) (Answer)
- [Pivot Example](#) (Answer)
- [Sales Trend by Day of Week](#) (Answer)
- [Customer Location](#) (Answer)
- [Customer Footprint and Sales by Region](#) (Answer)
- [Sales by Region, State and Year](#) (Answer)
- [Low Inventory](#) (Answer)
- [Monthly Department Sales Analysis](#) (Answer)
- [Sales for Last Month](#) (Answer)
- [Product & Department Sales Group Sum](#) (Answer)
- [Margin vs Sales Analysis](#) (Answer)
- [Sales by Age Group, Gender and Product Category](#) (Answer)
- [Racquet Sales](#) (Answer)
- [Sales by Quarter](#) (Answer)
- [Average Sales - Weekday vs Weekend](#) (Answer)
- [Vicky's Sales Data](#) (Answer)
- [Sales Per Customer for Outerwear by State](#) (Answer)

OK

5. You can also click the name of a worksheet or table and then click **Dependents**, to see a list of dependent objects with links.

The **Dependents** list shows the names of the dependent objects (worksheets and pinboards), and the columns they use from that source. You can use this information to determine the impact of changing the structure of the data source or to see how widely used it is. Click a dependent object to modify or delete it.

Delete a worksheet or table

WORKSHEET

Sporting Goods Retail Worksheet

ThoughtSPORT worksheet

Columns Schema Data **Dependents** Row Security SpotIQ Insights

COLUMN NAME	DEPENDENT NAME	TYPE
Age Group	Total Sales by Depar..	View
Date	Top 100 Products M..	View
Product Name	Top 100 Products M..	View
Department	Total Sales by Depar..	View
Sales	Total Sales by Depar..	View
Customer City	Customer Location	Answer
Customer Name	Customer Location	Answer

(showing rows 1-0 of 27)

THOUGHTSPORT

Understanding views

Summary: If you want to perform a search on top of another search, try saving your search as a view. Then, you can use the saved view as a data source for a new search.

Q Access: Only users with the **Can administer ThoughtSpot** or the **Can manage data** privilege can create views and link them.

A Important: Views do not support row level security (RLS), so all users of a view can see all the data it contains.

Introduction to views

You may have noticed that when you do a search on a data source, ThoughtSpot is only able to aggregate one column by one other column. Because of this, you may come across searches you can't do in one pass, because they are essentially nested queries. But you can create the equivalent of nested queries using a view, which is an answer that you have saved for the purpose of building other searches on top of it.

You can use a view just like any other data source. You can even link it to other sources by defining a relationship. When you save an answer as a view, and then use it as a source in a new search, it is similar to doing a nested query in SQL, only much easier to create.

View workflow

Suppose you created a search on the sales fact table that shows the top ten Sales Reps by revenue, for the first quarter. Then you want to do some further investigation on that subset of data, such as ranking them by how much they discounted a specific product based on data from the orders fact table. Unless you save your first answer as a view, and then search over that view, you cannot get your answers.

Here are the high-level steps for creating and using views:

1. Create the first search, and save it as a view [See page 497].
2. [Create relationships \[See page 450\]](#) or [define joins \[See page 191\]](#) to connect your view with any other data source.
3. Create a new search that includes your view and the other sources linked with it.
4. We recommend that you [create a new worksheet \[See page 461\]](#) that includes all these data sources.

Creating a worksheet makes it easier for people to search using your view and any related tables.

Best practices for using views

- When creating views, keep in mind the sizing recommendations for worksheets, for the final worksheet that you plan to use in modeling your data.
- To be able to join a view with a base table, your installation must be configured to allow this. The view cannot have more than 5 tables, and the number of rows in the view cannot exceed 10 million rows.
- The order of the objects being linked (joined) matters, because joins are directional. The table or view with the foreign key must be in the first (left) position. The table or view with the primary key must be in the second (right) position.
- For best performance, views should have 50 or fewer columns, and no more than 10 million rows. If your View has more than 10 million rows, consider materializing it. If your View has more than 40 million rows, consider sharding it.
- To improve performance, you can materialize the views.
- You can use an ETL (extract, transform, load) processes to circumvent these limitations.

Related Information

- [More view scenario examples \[See page 501\]](#)

- Save a search as a view [See page 497]
- Constraints [See page 191]
- Materialized views [See page 503]

Save a search as a view

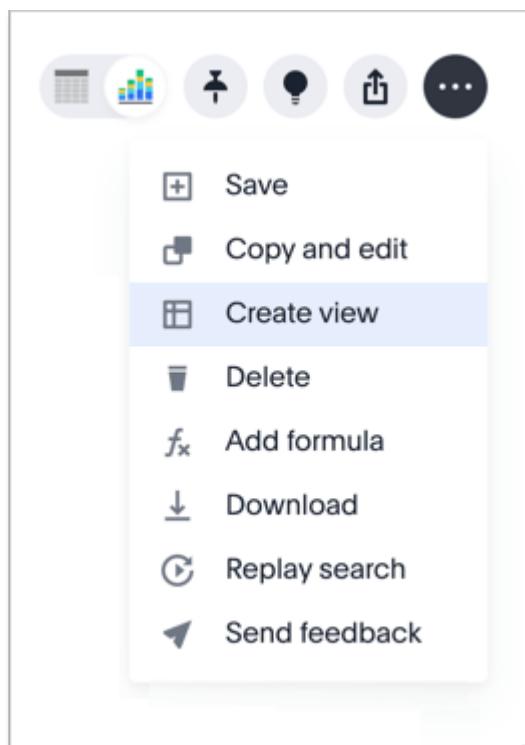
Summary: If you want to search on top of another search, try saving your search as a view. Then, you can use the saved view as a data source for a new search.

This procedure walks you through creating a view from a search. To create a view from a search:

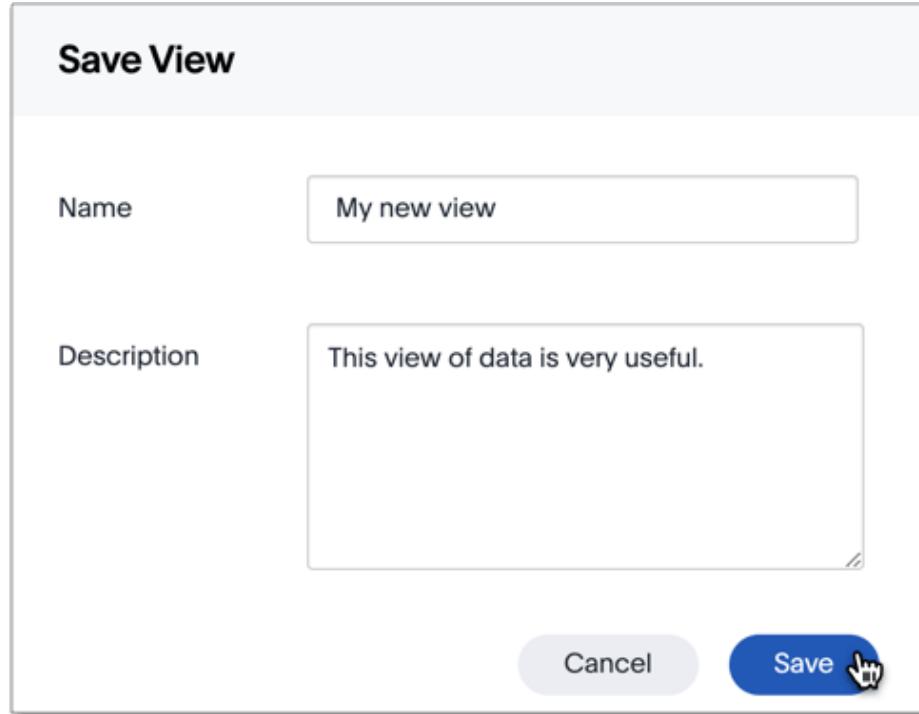
1. Start a new search, or edit an existing Answer, or visualization from a Pinboard.

Any filters or aggregations created during this search become part of the new view.

2. Make any changes to the visualization that you want in your saved view (change aggregation level, filters, etc.)
3. Click the **More** menu  (ellipses icon), and select **Create view**.



4. Give the view a name and save it.



5. [Link \[See page 450\]](#) your view to any other data source, or [define joins \[See page 191\]](#).

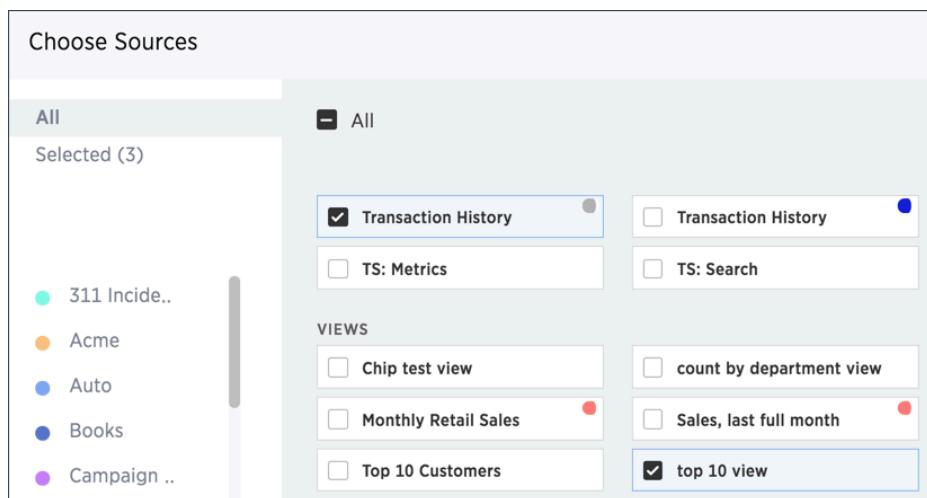
Your view is now saved. You may want to [materialize the view \[See page 503\]](#) to improve its performance when searching.

Create a search from a view

Summary: After creating a view, you can select it and search its data.

After creating a view and linking it to related data, you're ready to create your new search. To do a search on the view, along with any data sources you linked:

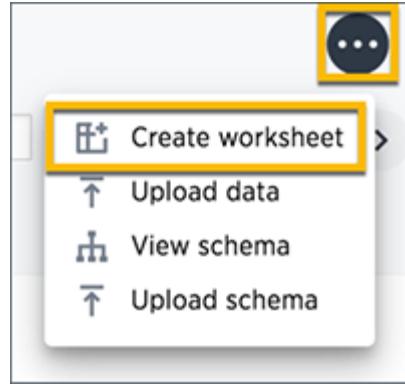
1. Click the search icon, and select **Data Source**. Choose your linked sources.



2. Do a search using columns from the linked sources.
3. Test the result, to make sure it's what you expect.

If your search shows no data found or doesn't look right to you, it is possible that one of the links between your sources was made on the incorrect column. Check the relationships you created and try linking using a different column, to see if that gives the expected search results.

4. When you have the expected answer, you can create a worksheet to make it easier for you and other people to use. To do this, click **Data**.
5. Click the ellipses icon , and select **Create worksheet**.



6. Choose your view and any other tables it is linked to using **Choose Sources**.
7. You may need to rename some columns appropriately for searching.
8. Save the worksheet and share it with the appropriate users and groups.

View example scenarios

Summary: It's not always easy to know when you need to use a view, but these are some common scenarios.

Here are three common examples of when you would want to use a view.

Example 1

Search a filtered subset

Let's say you have a table with all the sales transactions for each customer. You want to find high value customers (those who have spent a certain amount of money in the last year). Then you want to do some further analysis on these high value customers, like the count by region. Here you essentially need to sum up the sales per customer, filter by that sum to get the top customers, and then do a count, making it two levels of aggregation. A single search can do only one level of aggregation.

However, a search can calculate the initial sum and filter within it to get the highest value customers. Then you can save that result as a view. Joining the view to some other tables, like the customer dimension, allows you to do the count and other similar analyses using the view along with the other tables.

Example 2

Cohort analysis

This example is similar to the previous example, except that it involves a more complex filter, called a cohort. You might create a search to find customers who bought product A, but did not buy product B.

First, you would have to perform searches for total sales by customer for both A and B, and create views for both. Then join these two views back in an outer join looking for conditions where the A and B join values are null.

This example could also be solved in a single search by using [conditional formulas \[See page 0\]](#) to determine the desired group (or cohort) of customers as defined by their buying behavior, and then doing the analysis on this group. But the performance will be better if you do this using a view, and then [materialize the view \[See page 503\]](#).

Example 3

Combining aggregated sources

If you do a lot of your analysis in the aggregate, like at the quarterly level, you may be challenged when trying to combine this aggregated data from different sources. For example, you may want to combine three different quarterly views from SalesForce - Pipeline, Bookings, and Sales Activity to create a comprehensive analysis of each region's performance. Joining the quarterly views together in a worksheet enables you to do this type of analysis easily.

Example 4

Lightweight data transformation

You may have a table with separate columns for people's first name and last name. You could use a concat formula to create a single column with first name, a space, and last name. But using this formula could make performance slow, since that column would always be calculated on the fly when searching. In this case, creating a view using the concatenate formula and materializing it will improve performance when searching this data.

About materialized views

Summary: You can materialize a view to improve its performance.

Introduction to materialized views

Views can become slow as your data volume and the number of views stacked on top of one another grow. This happens because when doing a search, each view is computed on the fly. To improve the performance of a view, you can materialize it. Materializing pre-computes the view and stores it in memory, just like a table. You can expect a materialized view to be three to six times faster than a view that is not materialized, on average.

To materialize a view, you must have the [Can administer ThoughtSpot](#) privilege [See page 265].

Note: The number of materialized views you can create is limited to 50, because materialized views take up space in memory.

Row Level Security does not apply to materialized views. All users will see the same data when using the materialized view as a data source.

These are the operations you can do on a view:

- [Materialize a view](#) [See page 506] to improve performance,
- [Dematerialize a view](#) [See page 513] to save space, or
- [Refresh a view](#) [See page 515] to synchronize its data with the latest data load to the underlying tables.
- Check the status of a view.

After materializing, a view can have one of three statuses. You can see the status of a view by selecting **Data**, clicking on **Views**, and finding your view in the list. The status is shown next to the name of your view.

Materialization status

[See page 503]

A materialized view can be in one of the following states:

Status	Description	Action
Queued	View is waiting to be materialized.	Click Update status .
In Progress	View is being materialized.	Click Update status . Wait for materialization to finish.
Materialized	View has been materialized into memory.	You can now search the materialized view.
Stale	View has become out of sync with the data in the tables that make it up, usually due to an incremental data load.	Refresh the view by clicking Refresh Data* .
Unknown	View status could not be determined.	Click Update status .
Error	An error occurred.	Refresh the view by clicking Refresh Data* . Optionally run a report.

* You can refresh data only if you have the [Can administer ThoughtSpot](#) [See page 265] privilege.

Freshness of data

Your view can become stale, if it isn't refreshed when the data in the tables used to create it is updated.

There are two ways to refresh the data:

- [Manually refresh data in a view](#) [See page 515]
- [Refresh the view on a schedule](#) [See page 517]

Here are some conditions that could lead to a view becoming stale:

- Changes to the view definition
- Changes to the fields used in the view
- Changes to the join relationships used in the view
- Change to formulas in the view
- Change the datatypes in the underlying tables

Related Information

- [Understand views \[See page 494\]](#)
- [Save a search as a view \[See page 497\]](#)
- [Materialize a view \[See page 506\]](#)

Materialize a view

Summary: Learn how to materialize a view to improve its performance.

To improve the performance of a view, you can materialize it. This operation computes the data for the view and stores it in memory, just like a table. You can expect a 3-6 times performance improvement on a materialized view over a view that is not materialized. However, materialized views take up space in memory, so weigh the tradeoffs and benefits of materializing a view ahead of time.

Note: Row Level Security does not apply to materialized views. All users will see the same data when using the materialized view as a data source.

To materialize a view:

1. To find your view, click **Data** in the top menu.
2. Under **Data Objects** at the top of the page, choose **Views**.
3. Click the name of your view.
4. Click **Joins**.
5. If you don't have any joins, you may wish to add them now. See [add joins \[See page 450\]](#).
6. Under **Materialization**, click **Materialize**.

Materialization

When a view is materialized, all the data that makes up the view is stored in memory to improve query performance. [Learn more](#)

Status

Not Materialized

Materialize

7. Fill in the details for primary key [See page 191], sharding key, and number of shards [See page 196].

Materialize View

These settings are important for optimizing performance on a large materialized view. If you do not set them, the view will be replicated on every node of your cluster. [Learn more](#)

Primary Key (Optional)

Select keys

Sharding Key (Optional)

Select keys

Number of Shards (Optional)

Number of Shards

Cancel **Materialize**

Note: If you do not fill in the sharding details, the view is replicated on every node of your cluster. Follow the links in this step to learn more about sharding.

8. Click **Materialize**.
9. In the **Schedule data updates** dialog, decide if you want to configure the schedule to refresh materialized view at this time.

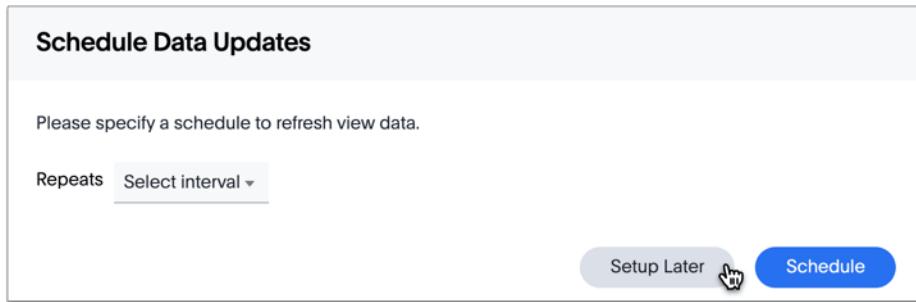
If you want to skip this step and perhaps schedule it later, click **Setup later**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats [Select interval ▾](#)

[Setup Later](#)  **Schedule**



10. To schedule a refresh for the materialized view, next to **Repeats**, select the **Time Interval**:
Daily, Weekly, or Monthly.,

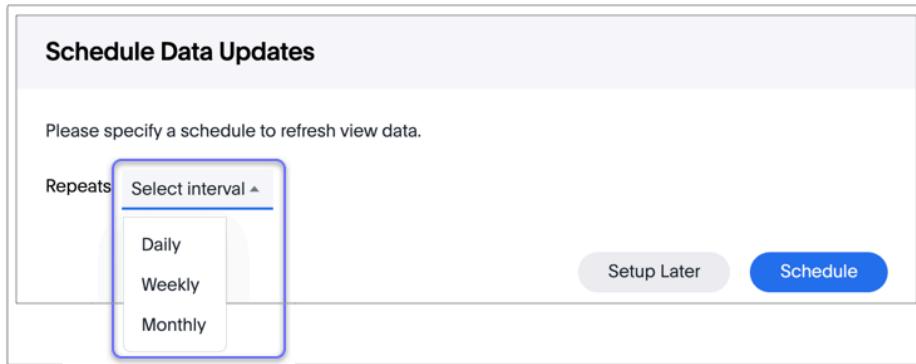
Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats [Select interval ▾](#)

Daily **Weekly** **Monthly**

[Setup Later](#)  **Schedule**



11. Supply the schedule details:

Daily

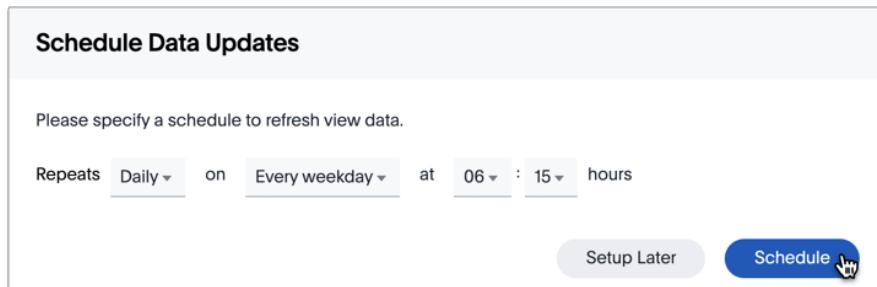
- From the **Select days** menu, select *Every weekday*, or *Every day*.
- Specify the hour of the day.
- Specify the minutes.
- Click **Schedule**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats [Daily ▾](#) on [Every weekday ▾](#) at [06 ▾](#) : [15 ▾](#) hours

[Setup Later](#)  **Schedule**



Weekly

- From the **Select weekdays** menu, select the days of the week, *Sunday Saturday*, for the refresh, and click **Done**.
- Specify the hour of the day.

- Specify the minutes.
- Click **Schedule**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats **Weekly** on **Monday + 2** at **06 : 15** hours

Setup Later **Schedule** 

Monthly

- From the **Select days** menu, select the day of the month, **1 ... 31**, for the refresh.
- Specify the hour of the day.
- Specify the minutes.
- Click **Schedule**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats **Weekly** on **Monday + 2** at **06 : 15** hours

Setup Later **Schedule** 

Note: Refresh works only if it is scheduled in the refresh window set for the cluster, with default: 8:00 PM - 4:00 AM. Only the start time of the refresh window is configurable using the flag `orion.materializationConfig.refreshWindowStartTime`. This flag can have values like 12:00PM, 01:00AM, and so on. For example, to set the cluster window from 2:00 AM to 10:00 AM, set the flag to `orion.materializationConfig.refreshWindowStartTime "02:00AM"`.

12. You can see the status of the materialization in progresses.

Materialization

When a view is materialized, all the data that makes up the view is stored in memory to improve query performance. [Learn more](#)

Status

! **In Progress** [Update Status](#)

Update Schedule: **None**

[Dematerialize](#) [Show Details](#)

The view cannot be dematerialized when it's queued or in progress.

13. Click **Update status** to refresh materialization status.
14. Note that when the materialization completes, the status changes to **Materialized**.

Materialization

When a view is materialized, all the data that makes up the view is stored in memory to improve query performance. [Learn more](#)

Status

✓ **Materialized**

Data Last Updated: 09/21/2020 09:35 AM
Update Schedule: **None**

[Dematerialize](#) [Show Details](#)

Materialized view without a schedule

Materialization

When a view is materialized, all the data that makes up the view is stored in memory to improve query performance. [Learn more](#)

Status

✓ **Materialized**

Data Last Updated: 09/21/2020 09:44 AM
Update Schedule: **Monthly**

[Dematerialize](#) [Show Details](#)

Materialized view with a monthly schedule

If after some time, the status has not changed to **Materialized**, check the [list of statuses \[See page 503\]](#) and corresponding actions to take.

When it says **Materialized** your materialized view is ready for search.

Related Information

- [Understand views \[See page 494\]](#)
- [Save a search as a view \[See page 497\]](#)
- [About materialized views \[See page 503\]](#)
- [Constraints \[See page 191\]](#)
- [Sharding \[See page 196\]](#)

Dematerialize a view

Summary: Learn how to dematerialize a view to free up space.

To free up space in your cluster, you can dematerialize views that were materialized but are no longer regularly used. This operation leaves the view intact, but removes the precomputed data that is stored in memory. After dematerializing a view, you can expect searches on the view to take longer, but you will have more space available in memory.

To dematerialize a view, follow these steps:

1. To find your view, click **Data** in the top menu.
2. Under **Data Objects** at the top of the page, choose **Views**.
3. Click the name of your view.
4. Click **Joins**.
5. Under **Materialization**, click **Dematerialize**.

The image contains two side-by-side screenshots of the ThoughtSpot interface, both titled "Materialization".

Left Screenshot (Materialized view without a schedule):

- Status:** Materialized (green button)
- Data Last Updated:** 09/21/2020 09:35 AM
- Update Schedule:** None
- Buttons:** Dematerialize (highlighted with a blue box and a cursor icon), Show Details

Right Screenshot (Materialized view with a monthly schedule):

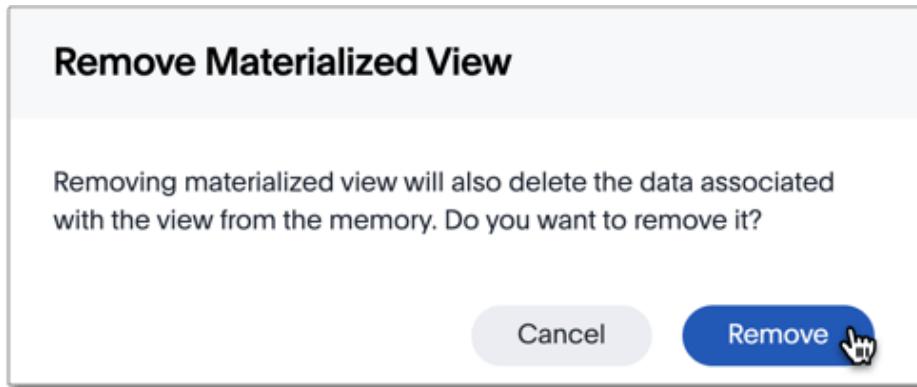
- Status:** Materialized (green button)
- Data Last Updated:** 09/21/2020 09:44 AM
- Update Schedule:** Monthly
- Buttons:** Dematerialize (highlighted with a blue box and a cursor icon), Show Details

Labels below the screenshots:

- Materialized view without a schedule
- Materialized view with a monthly schedule

6. In the confirmation window titled **Remove materialized view**, click **Remove** to confirm.

Note that this action removes the data from memory.



7. You can see the progress of dematerialization. Click **Update status** to refresh.

Remember that at this point, you can still search across your view, even if it is not materialized.

To remove the view altogether, you must delete it.

Related Information

- [Understand views \[See page 494\]](#)
- [Materialize a view \[See page 506\]](#)

Refresh a materialized view

Summary: You can refresh a view manually, so that its data matches the data in the underlying tables.

Materialized views can get out of sync with their underlying data. If the status of a view is **Stale**, you must refresh the view manually to sync it with the table.

Note that until it is refreshed, the view remains searchable.

To refresh a view manually, follow these steps:

1. To find your view, click **Data** in the top menu.
2. Under **Data Objects** at the top of the page, choose **Views**.
3. Click the name of your view.
4. Click **Joins**.
5. In the **Materialization** panel, notice that the **Status** is *Stale*.
6. Next to the *Stale* status, click **Refresh Data**.

Materialization

When a view is materialized, all the data that makes up the view is stored in memory to improve query performance. [Learn more](#)

Status

! **Stale** [Refresh Data](#) 🕒

Data Last Updated: 08/27/2019 01:23 PM
Update Schedule: **None**

[Dematerialize](#) [Show Details](#)

Related Information

- [Understand views \[See page 494\]](#)
- [About materialized views \[See page 503\]](#)

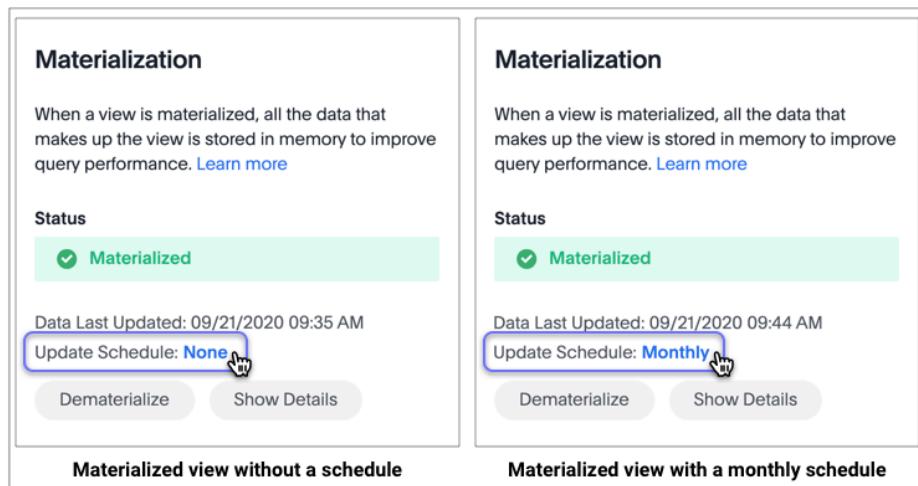
Create or change schedule for refreshing a materialized view

Summary: Learn how to create or change the schedule for data refresh in a materialized view.

To keep the data in a materialized view consistent and relevant, we recommend that you periodically refresh it with data from the parent table or tables. ThoughtSpot makes it easy, by letting you schedule regular refreshes at daily, weekly, or monthly intervals.

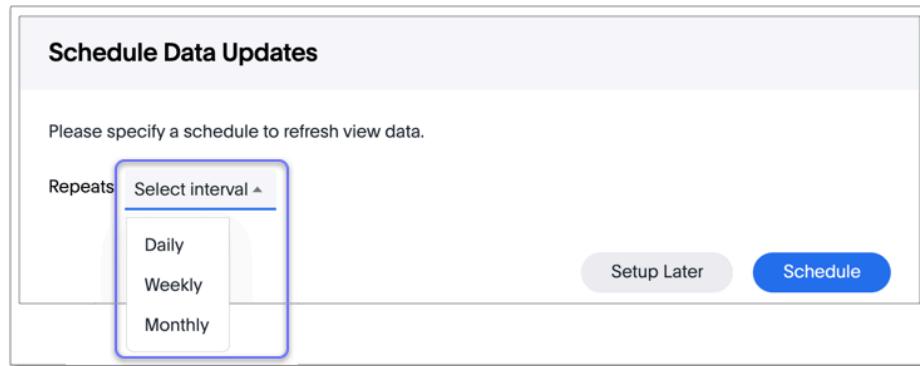
To schedule materialization of a view, follow these steps:

1. To find your view, click **Data** in the top menu.
2. Under **Data Objects** at the top of the page, choose **Views**.
3. Click the name of your view.
4. Click **Joins**.
5. Under **Materialization**, in the **Update schedule** section, either update an existing schedule, or create a new schedule:
 - To update an existing schedule, click *Daily*, *Weekly*, or *Monthly*.
 - To create a schedule, click *None*.



6. To schedule a refresh for the materialized view, next to **Repeats**, select the **Time Interval**:

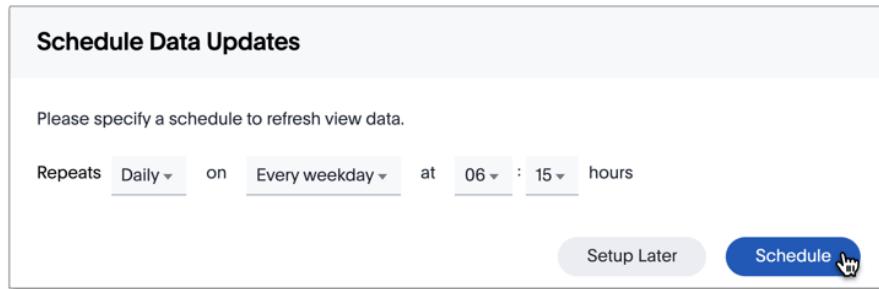
Daily, Weekly, or Monthly.,



7. Supply the schedule details:

Daily

- From the **Select days** menu, select *Every weekday*, or *Every day*.
- Specify the hour of the day.
- Specify the minutes.
- Click **Schedule**.



Weekly

- From the **Select weekdays** menu, select the days of the week, *Sunday Saturday*, for the refresh, and click **Done**.
- Specify the hour of the day.
- Specify the minutes.
- Click **Schedule**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats **Weekly** on **Monday + 2** at **06 : 15 hours**

Setup Later **Schedule**

Monthly

- From the **Select days** menu, select the day of the month, **1 ... 31**, for the refresh.
- Specify the hour of the day.
- Specify the minutes.
- Click **Schedule**.

Schedule Data Updates

Please specify a schedule to refresh view data.

Repeats **Weekly** on **Monday + 2** at **06 : 15 hours**

Setup Later **Schedule**

Note: Refresh works only if it is scheduled in the refresh window set for the cluster, with default: 8:00 PM - 4:00 AM. Only the start time of the refresh window is configurable using the flag `orion.materializationConfig.refreshWindowStartTime`. This flag can have values like `12:00PM`, `01:00AM`, and so on. For example, to set the cluster window from 2:00 AM to 10:00 AM, set the flag to `orion.materializationConfig.refreshWindowStartTime "02:00AM"`.

Related Information

- [Understand views \[See page 494\]](#)
- [Save a search as a view \[See page 497\]](#)
- [About materialized views \[See page 503\]](#)

Scriptability overview

Summary: Use ThoughtSpot's Scriptability to migrate or edit Worksheets, Pinboards, Answers, Views, and tables in a human-readable format.

ThoughtSpot developed its own scriptable approach for exporting, enhancing, and migrating Worksheets, Pinboards, Answers, Views, and tables.

You can model your data and build out sophisticated dashboards in your test environment, before deploying to all users.

The Scriptability [See page 521] feature supports several scenarios that you may encounter:

- **Migrating from a development environment to a production environment** by downloading files from the development cluster and uploading the same files into the production cluster.
- **Implementing metadata changes outside ThoughtSpot UI**, such as replacing the underlying tables for an object, or replacing a single column from one table with a column in another table.
- **Making bulk changes**, such as mass renaming of objects defined by Worksheets, and managing duplicates.
- **Reusing existing objects to build new objects**, such as building two very similar objects based on a similar, pre-existing object.

You can download these objects in the form of a scriptable file in the [ThoughtSpot Modeling Language \[See page 535\]](#), or TSL.

Use Scriptability to export Worksheets, Views, tables, Pinboards, and Answers in a human-readable format. Refer to [Migrate or restore objects \[See page 521\]](#).

Scriptability

Summary: Use Scriptability to export and import Worksheets, Views, Tables, Pinboards, and Answers in a human-readable format.

ThoughtSpot developed its own scriptable approach for exporting, enhancing, and migrating Worksheets, Views, Tables, Pinboards, and Answers.

You can model your data and build out sophisticated dashboards in your test environment, before deploying to all users.

The Scriptability feature supports several scenarios that you may encounter:

- **Migrating from a development environment to a production environment** by downloading files from the development cluster and uploading the same files into the production cluster
- **Implementing metadata changes outside ThoughtSpot UI**, such as replacing the underlying tables for an object, or replacing a single column from one table with a column in another table
- **Making bulk changes**, such as mass renaming of objects defined by Worksheets, and managing duplicates
- **Reusing existing objects to build new objects**, such as building two very similar objects based on a similar, pre-existing object.

How to use Scriptability

Depending on how you want to use Scriptability, there are several workflows you can follow:

1. **Edit and update an existing object in the same cluster:** You can either
 - [export \[See page 523\]](#) the object(s), edit the object(s) by modifying its [ThoughtSpot Scripting Language \[See page 535\]](#) (TSL) representation, and [import \[See page 528\]](#) the updated file(s) to modify the existing object *or*
 - edit the object(s) using the [in-app TSL editor \[See page 525\]](#) and publish the updated file(s).
2. **Migrate an existing object from one cluster to a new cluster:** [export \[See page 523\]](#) the object(s) and [import \[See page 531\]](#) the updated file(s) to the new cluster.
3. **Edit and migrate an existing object from one cluster to a new cluster:** You can either

- [export \[See page 523\]](#) the object(s), edit the object(s) by modifying its [ThoughtSpot Scripting Language \[See page 535\]](#) (TSL) representation, and [import \[See page 531\]](#) the updated file(s) to the new cluster or
- edit the object(s) using the [in-app TSL editor \[See page 525\]](#), publish the updated file(s), [export \[See page 523\]](#) the object(s), and [import \[See page 531\]](#) the updated file(s) to the new cluster. Note that this workflow changes the object(s) in both clusters.

Prerequisites

Refer to the following tables for required permissions for importing and exporting Pinboards, Answers, Worksheets, Tables, and Views.

Import

Import and create a new object without importing its dependents	Import and create a new object and its dependents	Import and update an existing object without dependents	Import and update an existing object with dependents
The dependents must already exist in the cluster. You must have view permissions for the first-level dependent. For example, if you import a Pinboard that is built on a Worksheet that is built on a table, you must have view permission for the Worksheet. When importing a new Worksheet or View, you must have the can manage data permission.	Can manage data , if the object or any of its dependents is a Worksheet or View.	Edit permission on the existing object. The dependents must already exist in the cluster. You must have view permissions for the first-level dependent. When importing a Worksheet or View, you must have the can manage data permission.	Edit permission on the existing object(s). Can manage data , if the object or any of its dependents is a Worksheet or View.

Export

Export with dependents	Export without dependents
View permission on the object and all dependents.	View permission on the object and its first-level dependents.

Note: If you have a permissions issue with a particular object when you export multiple objects, or an object and its dependents, the complete export does not fail. The individual object does not export, and you receive an error message about this failure in the `Manifest` file in the zip file.

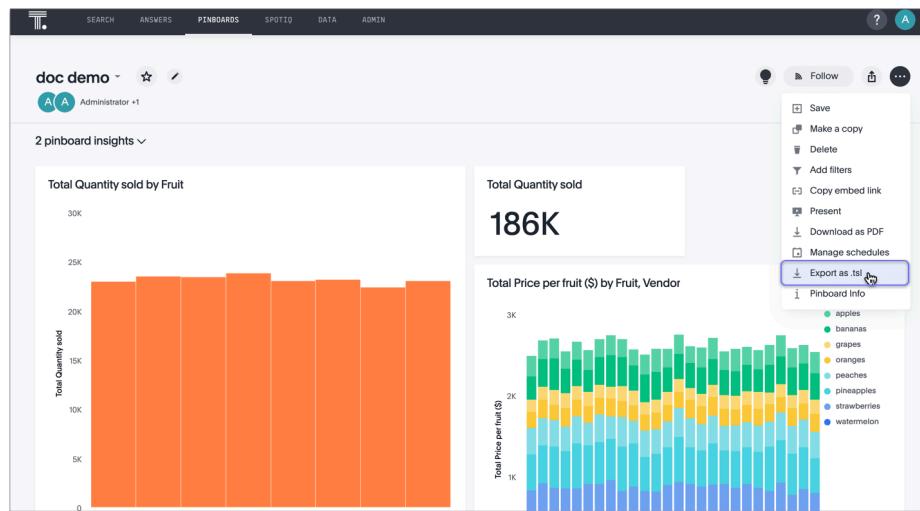
Export an object

You can export one object at a time [See page 523], or export more than one object as a zip file [See page 523], or SpotApp. The SpotApp contains a document called the `Manifest` file, which defines the objects you exported, and their underlying data sources.

Export one object

To export one object:

1. Navigate to the Pinboard, Answer, View, Table, or Worksheet you want to export.
2. Click the three-dot icon, and select **Export as TSL**.

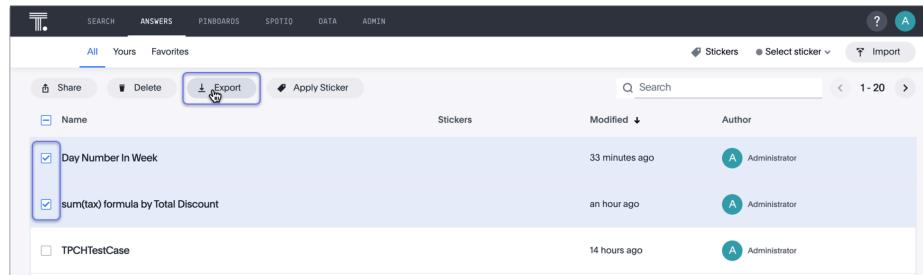


Export multiple objects

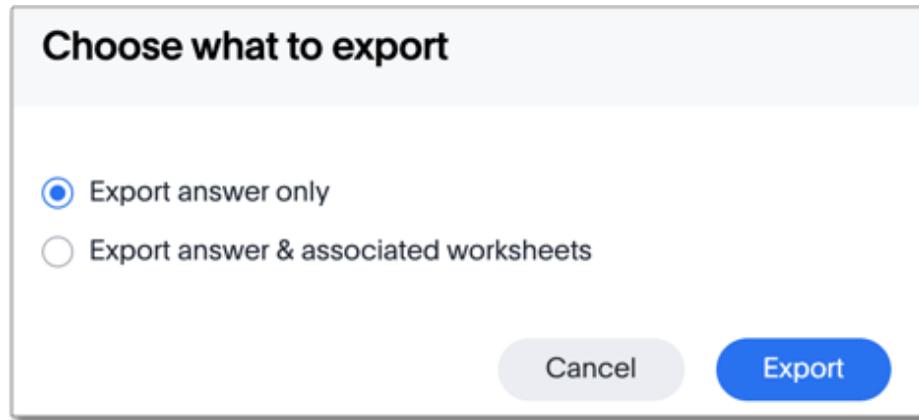
To export multiple objects at a time, follow these steps:

1. Navigate to the **Answers**, **Pinboards**, or **Data** page from the top navigation bar, depending on the objects you want to export.

2. Hover over the objects you want to export, and click the empty checkboxes that appear.
3. Select the **Export** button.



4. Choose whether to export only the objects, or the objects and their underlying data sources (Worksheets, Tables, and Views). If you export a table, you do not see this modal, since tables do not have any dependents.



5. Click **Export**.
6. Open the downloaded `TSL` zip file. The SpotApp zip file contains a document called the `Manifest` file, which defines the objects you exported, their underlying data sources, and any export errors. If an individual export fails, you can find an error message in the `Manifest` file. The zip file still exports, even if an individual object's export fails.

Edit the TSL file

You can edit the `TSL` file in one of two ways. You can [export \[See page 523\]](#) the object(s) and edit the file(s) in any text editor, before you import it. Or, you can use the [in-app `TSL` editor \[See page 0\]](#) to edit, validate, and publish the object(s). Refer to [ThoughtSpot Scripting Language \[See page 535\]](#) for information on syntax in the TSL files.

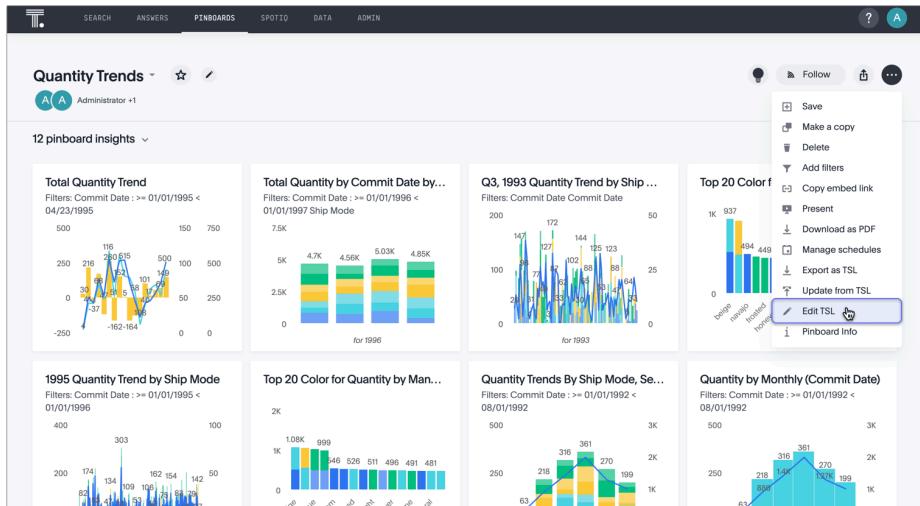
Edit, validate, and publish objects using the TSL editor

You can access the TSL editor from the object list page, or from the object itself. To edit and update multiple objects using the TSL editor, access it from the object list page.

To use the TSL editor, follow these steps:

1. Navigate to the **Answers**, **Pinboards**, or **Data** page from the top navigation bar, depending on the object you want to update.
2. Click the name of the object you want to edit, or select multiple objects by clicking on the checkboxes that appear when you hover over an object name.
3. From the object list page, select the **Edit TSL** button. From the object itself, select the ellipsis  (more options) menu in the upper-right side of the screen, and select **Edit TSL**.

The screenshot shows a list of TSL files in the ThoughtSpot interface. The top navigation bar includes SEARCH, ANSWERS, PINBOARDS, SPOTIQ, DATA, and ADMIN. Below the navigation is a toolbar with Share, Delete, Apply Sticker, Export, and Edit TSL (which is highlighted with a red box). A search bar and a page number indicator (1-20) are also present. The main area displays a table with columns for Name, Stickers, Modified (sorted by date), and Author. Several TSL files are listed, including "Agency Name, Employee Address, Employee First Name" (modified 51 minutes ago, author Administrator), "Basic Answer1" (modified 4 hours ago, author Administrator), "Average Revenue by Part" (modified 4 hours ago, author Administrator), "Brand Revenue" (modified 4 hours ago, author Administrator), "Average Revenue by Part" (modified 12 hours ago, author Administrator), "RandomAnswer" (modified 16 hours ago, author Administrator), "Average Revenue by Part" (modified 6 months ago, author Guest 1), and "Basic Answer1" (modified 6 months ago, author Administrator).



4. The TSL editor opens. Edit the TSL file(s), using the syntax specified in [ThoughtSpot Scripting Language \[See page 535\]](#).

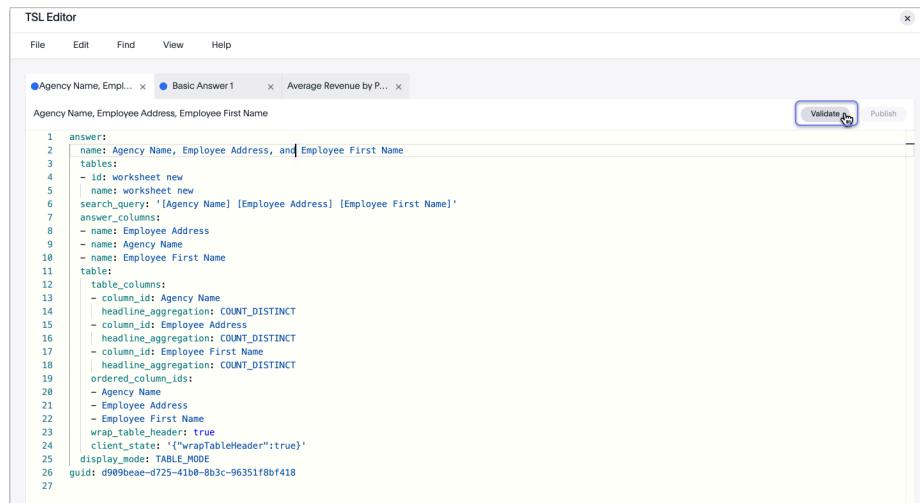
The TSL editor has the following functions under the top menu:

- **File:** Validate, Publish, and Exit editor. You can also validate and publish using the **validate** and **publish** buttons at the top right of the editor. You can also exit the editor using the X button at the top right corner. The system warns you if you try to exit with unsaved changes.
- **Edit:** Undo, Redo, Cut, Copy, Select all, Fold, Fold all, Unfold, Unfold all, and Go to line. The **Fold** option compresses the lines in the file so you only see the first line of a section. **Go to line** opens a dialog box, where you can type in the number of the

line you would like to go to. This is useful for long TSL files.

- **Find:** Find and Find and replace. This functionality allows you to easily find words or parameters in the TSL file. You can also click on a word or parameter in the TSL editor, and the editor highlights all instances of that word.
- **View:** Show/Hide errors, Show line numbers, and Hide line numbers. **Show/Hide errors** toggles the **Errors** sidebar on and off. The **Errors** sidebar does not appear until after you Validate a file, if there are errors in it.
- **Help:** Documentation. This links to the [ThoughtSpot Scripting Language \[See page 535\]](#) documentation.

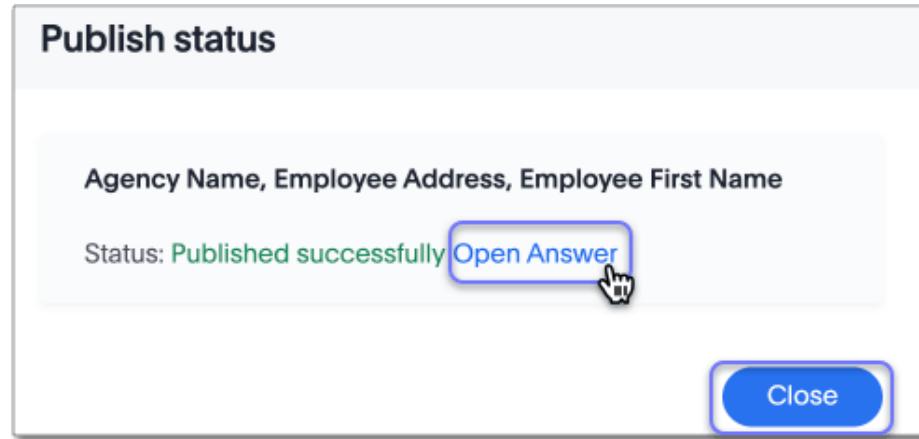
5. When you finish editing the TSL file(s), select **Validate** in the top right corner. You must validate each file individually. A blue dot appears next to any file that contains changes.



6. If you constructed the file(s) correctly, a green check mark appears next to the name of the file. If you did not construct the file correctly, a red bar appears near the top of the screen, telling you that ThoughtSpot found errors in one or more files. Click **Show errors** to see the errors listed in the **Errors** sidebar.



7. After validating, select **Publish** in the top right corner, next to **Validate**. You must publish each file individually.
8. The system displays a **Publish status** dialog box. You can select **Open [object]** to open the object you just published in a new tab, or click **Close** to return to the TSL editor.



Update an object

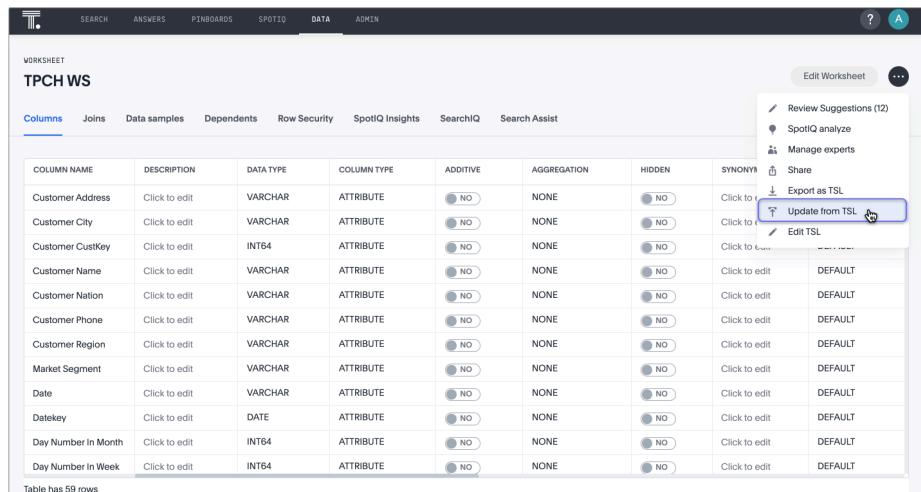
You can overwrite an existing Worksheet, View, Table, Answer, or Pinboard, by downloading the `TSL` file, making any necessary changes, and then re-uploading the `TSL` file. To update SpotApps, or collections of objects packaged together as a zip file, refer to [SpotApps \[See page 556\]](#).

You can also update an object using the [TSL editor \[See page 525\]](#).

To update an existing object by downloading the TSL file and modifying it, follow these steps. In this case, we are updating a single Worksheet. You can update multiple objects at once by uploading them in .zip file format.

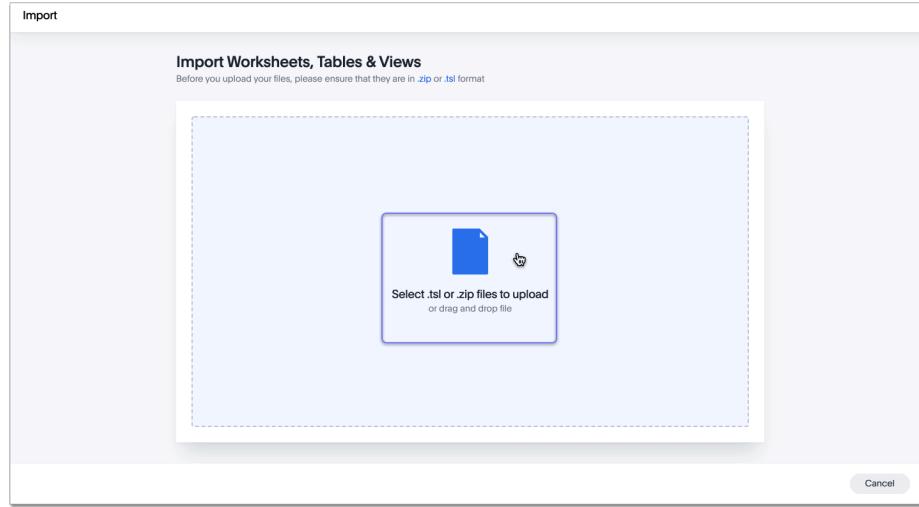
1. Export the object [See page 523] you want to update, as in steps 1 to 5 of the **Export an Object** section above.
2. Edit the file in a text editor.
3. Navigate to the **Answers**, **Pinboards**, or **Data** page from the top navigation bar, depending on the object you want to update.
4. Click the name of the object you want to edit.
5. Click the ellipsis  (more options) menu in the upper-right side of the screen.
6. Select **Update from TSL**.

Here, we are uploading the edited *TCPH WS* worksheet.

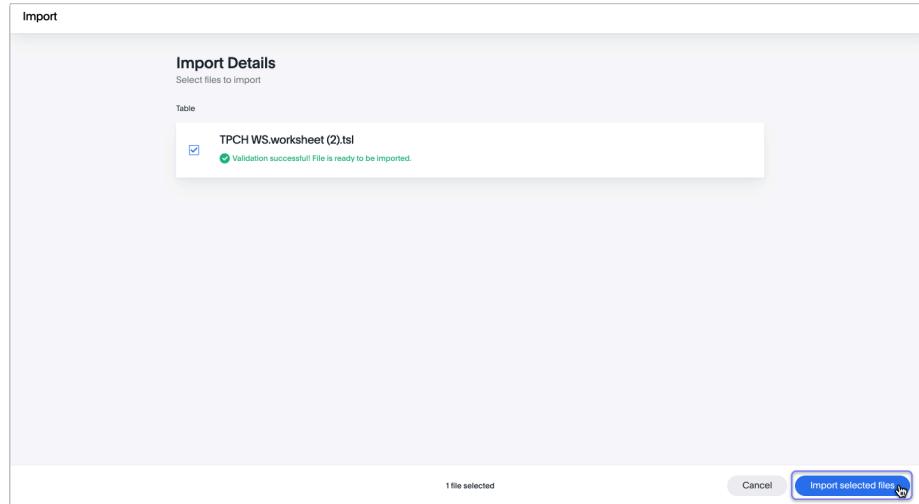


The screenshot shows the ThoughtSpot Worksheet interface for a 'TPCH WS' worksheet. The top navigation bar includes 'SEARCH', 'ANSWERS', 'PINBOARDS', 'SPOTIQ', 'DATA', and 'ADMIN'. Below the navigation is a toolbar with 'Columns', 'Joins', 'Data samples', 'Dependents', 'Row Security', 'SpottIQ Insights', 'SearchIQ', and 'Search Assist'. A context menu is open on the right, listing 'Edit Worksheet', 'Review Suggestions (12)', 'SpotIQ analyze', 'Manage experts', 'Share', 'Export as TSL', and 'Update from TSL' (which is highlighted with a blue border). The main area displays a table with 12 columns and 59 rows, with various column settings like 'ATTRIBUTE', 'VARCHAR', and 'INT64' visible.

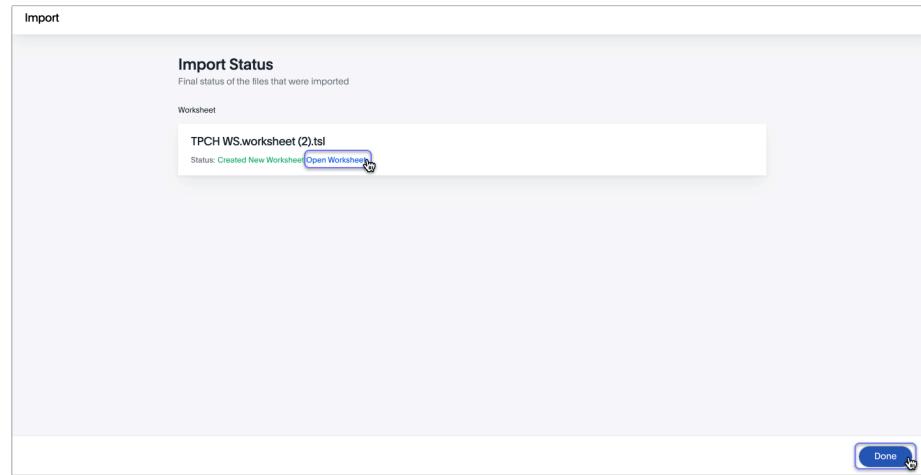
7. In the **Import** interface, click **Select .TSL or .zip files to upload**.



8. In your file system, find and select the `TSL` file you edited.
9. If you constructed the file correctly, the **Import** interface displays a *Validation successful* message. You can now import the file.
10. If you uploaded a `.zip` file with multiple objects, you can unselect any files in the `.zip` file you do not want to upload.
11. Click **Import selected files**.



12. The **Import Status** screen displays the status of the objects you imported. You can open the object(s) that you imported, or click **Done** to return to the main object page.



Migrate an object

To migrate an Answer, Pinboard, View, or Worksheet from one cluster to another, follow these steps. To migrate SpotApps, or collections of objects packaged together as a zip file, refer to [SpotApps \[See page 556\]](#). Note that you cannot create a new Table using Scriptability. You can only update existing Tables.

1. [Export the object \[See page 523\]](#) you want to move, as in steps 1 to 5 of the **Export an Object** section above.

The object remains on the original cluster as well, unless you delete it.

2. Navigate to the cluster you want to add the object to.
3. Click **Answers**, **Pinboards**, or **Data** on the top navigation bar, depending on the objects you want to migrate.
4. To upload a Worksheet or View, click the More icon in the upper-right side of the screen. Then, select **Import TSL**.

The screenshot shows the ThoughtSpot interface for managing data objects. The top navigation bar includes SEARCH, ANSWERS, PINBOARDS, DATA, and ADMIN. Under the DATA tab, there are sub-options for Data Objects, Embrace, Settings, and Usage. The main content area displays a table of worksheets, each with columns for Name, Source, Stickers, Materialize Status, Modified, and Author. A row for 'Retail - Apparel' has an 'i' icon. In the top right, there are buttons for Create worksheet, Import TSL (which is highlighted with a blue border), and View Schema.

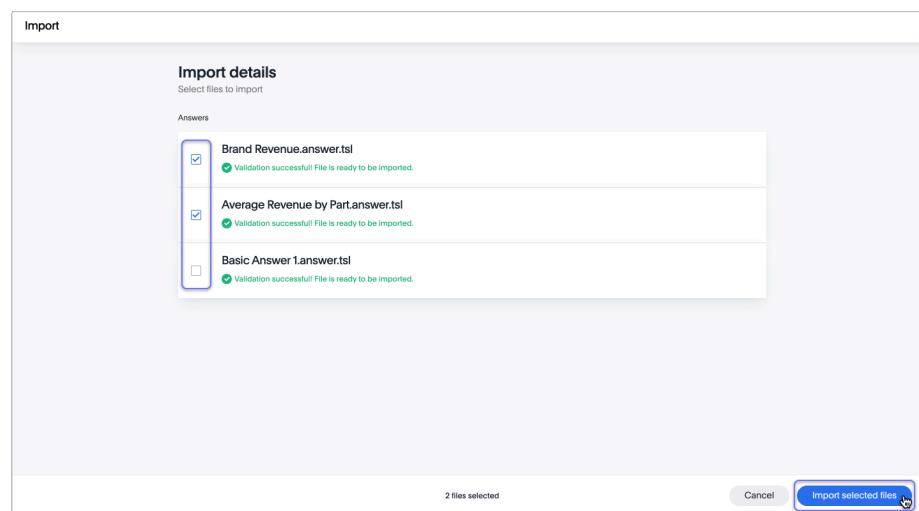
- To upload a Pinboard or Answer, click the **Import TSL** button in the upper-right side of the screen.

The screenshot shows the ThoughtSpot interface for managing pinboards. The top navigation bar includes SEARCH, ANSWERS, PINBOARDS, SPOTIQ, DATA, and ADMIN. Under the PINBOARDS tab, there are sub-options for All, Yours, and Favorites. The main content area displays a table of pinboards, each with columns for Name, Stickers, Modified, and Author. A search bar at the top right shows 'Q Search' and a page number '1 - 20'. A 'Pinboard' button is also visible in the top right.

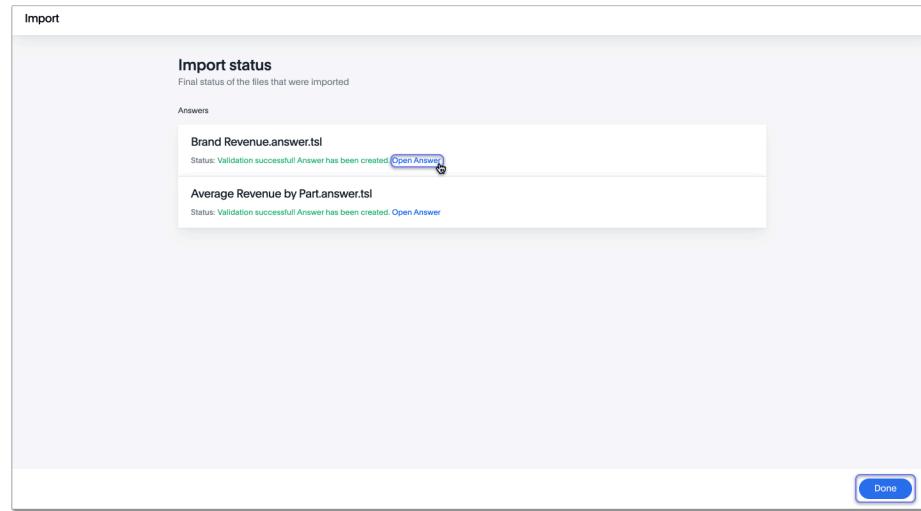
- In the **Import** interface, click **Select .TSL or .zip files to upload**.

The screenshot shows the ThoughtSpot import interface. The title bar says 'Import'. The main area is titled 'Import your answers' with a note: 'Before you upload your file below, make sure it is in .tsl'. It features a large dashed rectangular area for file selection, with a central button labeled 'Select .tsl file to upload or drag and drop file'. A 'Cancel' button is located at the bottom right.

7. In your file system, find and select the `TSL` file. The file uploads automatically.
8. If you constructed the file correctly, the **Import** interface displays a *Validation successful* message. You can now import the file.
9. If you uploaded a `.zip` file with multiple objects, you can unselect any files in the `.zip` file you do not want to upload. Here, we only want to import **Brand Revenue** and **Average Revenue by Part**, not **Basic Answer 1**.
10. Click **Import selected files**.



11. The **Import Status** screen displays the status of the objects you imported. You can open the object(s) that you imported, or click **Done** to return to the main object page.



Limitations of working with TSL files

There are certain limitations to the changes you can apply by editing a Worksheet, Answer, Table, View, or Pinboard through TSL.

- Formulas and columns can either have a new name, or a new expression. You cannot change both, unless migrating or updating the worksheet two times.
- It is not possible to reverse the join direction in the TSL script.
- You cannot create new tables using Scriptability. You can only update existing tables.
- You can only change logical tables using Scriptability. You cannot change the physical version of the table that exists in a database. When you change the `column_name`, for example, the name changes in the application, but not in the physical table in the database.
- You cannot import manually compressed .zip files. You can only import .zip files that you exported from ThoughtSpot: either an object and its associated data sources, or multiple objects of the same type that you exported from the object list page.

Related information

- [ThoughtSpot Scripting Language \[See page 535\]](#)

ThoughtSpot Scripting Language

Summary: Use ThoughtSpot Scripting Language to modify a Worksheet, View, table, Pinboard, or Answer in a flat-file format. Then you can migrate the object to a different cluster, or restore it to the same cluster.

To work with Scriptable Worksheets [See page 535], Views [See page 538], tables [See page 541], Answers [See page 543], and Pinboards [See page 546] in ThoughtSpot, you can download these objects to a flat file in `TSL` format, modify it, and subsequently upload this file either to the same cluster, or to a different cluster. To learn how to export, change, and update Worksheets, Views, Tables, Answers, and Pinboards, see [Scriptability \[See page 521\]](#).

Syntax of the Worksheet TSL file

The `TSL` file for Scriptable Worksheets has a specific syntax.

See the [Parameters \[See page 547\]](#) section for details about the keywords used in this example.

You may not see each of these parameters in your own TSL files, depending on whether each variable is explicitly defined. For example, if you do not have any filters on your Worksheet, the `filters` parameter does not appear. You can add that variable to the TSL file to specify filters for your Worksheet.

```

worksheet [See page 0]:
  name [See page 0]: <worksheet_name>
  description [See page 0]:
    This is a multi-line description of the worksheet
    Description line 2
  tables [See page 0]:
  - name [See page 0]: <table_name_1>
  - [alias [See page 0]] : <table_alias>
  - [fqn [See page 0]] : <GUID_of_table_name>
  - name [See page 0]: <table_name_2>
  - name [See page 0]: <table_name_3>
  joins:
  - name [See page 0]: <join_name_1>
    source [See page 0]: <source_table_name>
    destination [See page 0]: <destination_table_name>
    type [See page 0]: [RIGHT_OUTER | LEFT_OUTER | INNER | OUTE
R]
  on [See page 0]: <on_string>
  is_one_to_one [See page 0]: [ false | true ]
  -
  table_paths [See page 0]:
  - id [See page 0]: <table_path_name_1>
    table [See page 0]: <table_name_1>
    join_path [See page 0]:
    - join [See page 0]:
      - <join_name_1>
  - id [See page 0]: <table_path_name_2>
    table [See page 0]: <table_name_2>
    join_path [See page 0]:
    - {}
  - id [See page 0]: <table_path_name_3>
    table [See page 0]: <table_name_3>
    join_path [See page 0]:
    - join [See page 0]:
      - <join_name_1>
    - join [See page 0]:
      - <join_name_2>
      - <join_name_3>
    - join [See page 0]:
      - <join_name_4>
      - <join_name_5>
      - <join_name_6>
  formulas [See page 0]:
  - name [See page 0]: <formula_name_1>
  expr [See page 0]: <formula_definition_1>

```

```

[id]: <optional_unique_identifier>
- name [See page 0]: <formula_name_2>
  expr [See page 0]: <formula_definition_2>
- name [See page 0]: <formula_name_3>
  expr [See page 0]: <formula_definition_3>
filters [See page 0]:
- column [See page 0]: <filtered_column_name_1>
  oper [See page 0]: <filter_operator>
  values [See page 0]: <filtered_values>
  - value 1
  - value 2
  - value n
- column [See page 0]: <filtered_column_name_2>
worksheet_columns [See page 0]:
- name [See page 0]: <column_name_1>
  description [See page 0]: <optional_column_description>
  column_id [See page 0]: <column_id_1>
properties [See page 0]:
  column_type [See page 0]: [ MEASURE | ATTRIBUTE ]
  aggregation [See page 0]: [ SUM | COUNT | AVERAGE | MAX
| MIN | COUNT_DISTINCT | NONE | STD_DEVIATION | VA
RIANCE]
  index_type [See page 0]: [ DONT_INDEX | DEFAULT | PREFI
X_ONLY |
  PREFIX_AND_SUBSTRING | PREFIX_AND_WORD_SUBS
TRING ]
  index_priority [See page 0]: <index_priority>
synonyms [See page 0] :
  <synonym_1>
  <synonym_2>
is_attribution_dimension [See page 0] : [true | false]
is_additive [See page 0] : [ true | false ]
calendar [See page 0] : [ default | calendar_name ]
format_pattern [See page 0] : <format_pattern_string>
currency_type [See page 0] :
  is_browser : true
  OR
  column : <column_name>
  OR
  iso_code : <valid_ISO_code>
is_hidden [See page 0]: [ true | false ]
geo_config [See page 0] :
  latitude : true
  OR

```

```
longitude : true
    OR
country : true
    OR
region_name:
- country : <name_supported_country>
- region_name : <region_name_in_UI>
spotiq_preference [See page 0]: <spotiq_preference_string>
g>
    search_iq_preferred [See page 0]: [ true | false ]
    name [See page 0]: <column_name_2>
    description [See page 0]: <column_description>
    column_id [See page 0]: <column_id_2>
    ...
properties [See page 0]:
    is_bypass_rls [See page 0]: [ true | false ]
        join_progressive [See page 0]: [ true | false ]
    guid [See page 0]: <worksheet_guid>
```

Syntax of the View TSL file

The `TSL` file for Scriptable Views has a specific syntax.

See the [Parameters \[See page 547\]](#) section for details about the keywords used in this example.

You may not see each of these parameters in your own TSL files, depending on whether each variable is explicitly defined. For example, if you do not have a description for your View, the `description` parameter does not appear. You can add that variable to the TSL file to specify a description for your View.

```

view [See page 0]:
  name [See page 0]: <view_name>
  description [See page 0]:
    This is a multi-line description of the View.
    Description line 2
  tables [See page 0]:
    identity [See page 0]:
      - id [See page 0]: <table_id_1>
      - name [See page 0]: <table_name_1>
      - fqn [See page 0]: <table_fqn_1>
    identity [See page 0]:
      - id [See page 0]: <table_id_n>
      - name [See page 0]: <table_name_n>
      - fqn [See page 0]: <table_fqn_n>
  joins [See page 0]:
    - name [See page 0]: <join_name_1>
      source [See page 0]: <source_table_name>
      destination [See page 0]: <destination_table_name>
      type [See page 0]: [RIGHT_OUTER | LEFT_OUTER | INNER | OUTE
R]
      on [See page 0]: <on_string>
      is_one_to_one [See page 0]: [ false | true ]
  table_paths [See page 0]:
    - id [See page 0]: <table_path_name_1>
      table [See page 0]: <table_name_1>
      join_path [See page 0]:
        - {}
  formulas [See page 0]:
    - id [See page 0]: <formula_id_1>
      name [See page 0]: <formula_name_1>
      expr [See page 0]: <formula_definition_1>
      properties [See page 0]: <formula_properties_1>
        column_type [See page 0]: [ MEASURE | ATTRIBUTE ]
        data_type [See page 0]: [ BOOL | VARCHAR | DOUBLE | FLOA
T | INT | BIGINT | DATE | DATETIME | TIMESTAMP | TIME ]
        aggregation [See page 0]: [ SUM | COUNT | AVERAGE | MAX
| MIN |
          COUNT_DISTINCT | NONE | STD_DEVIATION
| VARIANCE]
      - id [See page 0]: <formula_id_n>
        name [See page 0]: <formula_name_n>
        expr [See page 0]: <formula_definition_n>
        properties [See page 0]: <formula_properties_n>
  filters [See page 0]:
    - column [See page 0]: <filtered_column_name_1>

```

```

oper [See page 0]: <filter_operator>
values [See page 0]: <filtered_values>
- value 1
- value 2
- column [See page 0]: <filtered_column_name_n>
view_columns [See page 0]:
- name [See page 0]: <column_name_1>
description [See page 0]: <optional_column_description>
column_id [See page 0]: <column_id_1>
phrase [See page 0]: <phrase_string_1>
properties [See page 0]:
    column_type [See page 0]: [ MEASURE | ATTRIBUTE ]
    aggregation [See page 0]: [ SUM | COUNT | AVERAGE | MAX
| MIN |
        COUNT_DISTINCT | NONE | STD_DEVIATION | VA
RIANCE]
    index_type [See page 0]: [ DONT_INDEX | DEFAULT | PREFI
X_ONLY |
        PREFIX_AND_SUBSTRING | PREFIX_AND_WORD_SUBS
TRING ]
        index_priority [See page 0]: <index_priority>
synonyms [See page 0] :
    <synonym_1>
    <synonym_2>
is_attribution_dimension [See page 0] : [true | false]
is_additive [See page 0] : [ true | false ]
calendar [See page 0] : [ default | calendar_name ]
format_pattern [See page 0] : <format_pattern_string>
currency_type [See page 0] :
    is_browser : true
    OR
    column : <column_name>
    OR
    iso_code : <valid_ISO_code>
is_hidden [See page 0]: [ true | false ]
geo_config [See page 0] :
    latitude : true
    OR
    longitude : true
    OR
    country : true
    OR
    region_name:
        - country : <name_supported_country>
        - region_name : <region_name_in_UI>

```

```
    spotiq_preference [See page 0]: <spotiq_preference_string>
    search_iq_preferred [See page 0]: [ true | false ]
    name [See page 0]: <column_name_2>
    description [See page 0]: <column_description>
    column_id [See page 0]: <column_id_2>
    ...
    query [See page 0]: <query_string>
    guid [See page 0]: <view_guid>
```

Syntax of the table TSL file

The `TSL` file for scriptable tables has a specific syntax.

See the [Parameters \[See page 547\]](#) section for details about the keywords used in this example.

You may not see each of these parameters in your own TSL files, depending on whether each variable is explicitly defined. For example, if you did not define an `index_priority` for your table, the `index_priority` parameter does not appear. You can add that variable to the TSL file to specify an index priority for the table.

```

table [See page 0]:
  name [See page 0]: <table_name>
  db [See page 0]: <database_name>
  schema [See page 0]: <schema_name>
  db_table [See page 0]: <database_table_name>
  connection [See page 0]:
    name [See page 0]: <connection_name>
    type [See page 0]: <connection_type>
  columns [See page 0]:
    - name [See page 0]: <column_name_1>
      db_column_name [See page 0]: <database_column_name>
      is_primary_key [See page 0]: [true | false]
      is_foreign_key [See page 0]: [true | false]
      data_type [See page 0]: [ BOOL | VARCHAR | DOUBLE | FLOAT
      | INT32 | INT64 | DATE | TIME ]
      properties [See page 0]:
        column_type [See page 0]: [ MEASURE | ATTRIBUTE ]
        aggregation [See page 0]: [ SUM | COUNT | AVERAGE | MAX
        | MIN |
          COUNT_DISTINCT | NONE | STD_DEVIATION | VA
          RIANCE]
        index_type [See page 0]: [ DONT_INDEX | DEFAULT | PREFI
          X_ONLY |
          PREFIX_AND_SUBSTRING | PREFIX_AND_WORD_SUBS
          TRING ]
        index_priority [See page 0]: <index_priority>
      synonyms [See page 0] :
        <synonym_1>
        <synonym_2>
        is_attribution_dimension [See page 0] : [true | false]
        is_additive [See page 0] : [ true | false ]
        calendar [See page 0] : [ default | calendar_name ]
        format_pattern [See page 0] : <format_pattern_string>
        currency_type [See page 0] :
          is_browser : true
          OR
          column : <column_name>
          OR
          iso_code : <valid_ISO_code>
        is_hidden [See page 0]: [ true | false ]
        geo_config [See page 0] :
          latitude : true
          OR
          longitude : true
          OR

```

```
country : true
OR
region_name:
- country : <name_supported_country>
- region_name : <region_name_in_UI>
spotiq_preference [See page 0]: <spotiq_preference_string>
search_iq_preferred [See page 0]: [ true | false ]
- name [See page 0]: <column_name_2>
- name [See page 0]: <column_name_n>
guid [See page 0]: <table_guid>
```

Syntax of the Answer TSL file

The `TSL` file for Scriptable Answers has a specific syntax.

See the [Parameters \[See page 547\]](#) section for details about the keywords used in this example.

You may not see each of these parameters in your own TSL files, depending on whether each variable is explicitly defined. For example, if you did not define any conditional formatting, the `conditional_formatting` variable does not appear. You can add that variable in the TSL file to specify conditional formatting.

```

answer [See page 0]:
  name [See page 0]: <answer_name>
  description [See page 0]:
    This is a multi-line description of the answer
    Description line 2
  tables [See page 0]:
    - id [See page 0]: <table_id>
    - name [See page 0]: <table_name_1>
    - [alias [See page 0]] : <optional_table_alias>
    - [fqn [See page 0]] : <optional_GUID_of_table_name>
  joins [See page 0]:
    - name [See page 0]: <join_name_1>
      source [See page 0]: <source_table_name>
      destination [See page 0]: <destination_table_name>
      type [See page 0]: [RIGHT_OUTER | LEFT_OUTER | INNER | OUTE
R]
      on [See page 0]: <on_string>
      is_one_to_one [See page 0]: [ false | true ]
    - ...
  table_paths [See page 0]:
    - id [See page 0]: <table_path_name_1>
      table [See page 0]: <table_name_1>
      join_path [See page 0]:
        - {}
  formulas [See page 0]:
    - id [See page 0]: <formula_id_1>
      name [See page 0]: <formula_name_1>
      expr [See page 0]: <formula_definition_1>
      properties [See page 0]: <formula_properties_1>
        column_type [See page 0]: [ MEASURE | ATTRIBUTE ]
        data_type [See page 0]: [ BOOL | VARCHAR | DOUBLE | FLOA
T | INT | BIGINT | DATE | DATETIME | TIMESTAMP | TIME ]
        aggregation [See page 0]: [ SUM | COUNT | AVERAGE | MAX
| MIN |
          COUNT_DISTINCT | NONE | STD_DEVIATION | VA
RIANCE]
    - id [See page 0]: <formula_id_2>
      name [See page 0]: <formula_name_2>
      expr [See page 0]: <formula_definition_2>
      properties [See page 0]: <formula_properties_2>
    - id [See page 0]: <formula_id_3>
      name [See page 0]: <formula_name_3>
      expr [See page 0]: <formula_definition_3>
      properties [See page 0]: <formula_properties_3>
  search_query [See page 0]: <search_query_string>

```

```

answer_columns [See page 0]:
- id [See page 0]: <column_id_1>
  name [See page 0]: <column_name_1>
  custom_name [See page 0]: <custom_name_1>
- name [See page 0]: <column_name_2>
table [See page 0]:
  table_columns [See page 0]:
    - column_id [See page 0]: <column_id_1>
      conditional_formatting [See page 0]:
        - range [See page 0]:
          min [See page 0]: <conditional_formatting_minimum>
          max [See page 0]: <conditional_formatting_maximum>
        - rule [See page 0]: <conditional_formatting_rule_1>
          range [See page 0]:
            min [See page 0]: <conditional_formatting_minimum>
            max [See page 0]: <conditional_formatting_maximum>
            color [See page 0]: <color_string>
            plotAsBand [See page 0]: [ true | false ]
        - rule [See page 0]: <conditional_formatting_rule_2>
          show_headline [See page 0]: [ true | false ]
          headline_aggregation [See page 0]: <headline_aggregation_string>
- column_id [See page 0]: <column_id_2>
  ordered_column_ids [See page 0]:
    - column_id [See page 0]: <column_id_1>
    - column_id [See page 0]: <column_id_2>
    client_state [See page 0]: <client_state_string>
chart [See page 0]:
  type [See page 0]: <chart_type>
  chart_columns [See page 0]: <chart_column_1>
  - column_id [See page 0]: <column_id_1>
    conditional_formatting [See page 0]:
      - rule [See page 0]: <conditional_formatting_rule_1>
        range [See page 0]:
          min [See page 0]: <conditional_formatting_minimum>
          max [See page 0]: <conditional_formatting_maximum>
          color [See page 0]: <color_string>
          plotAsBand [See page 0]: [ true | false ]
      - rule [See page 0]: <conditional_formatting_rule_2>
    - column_id [See page 0]: <column_id_2>
      axis_configs [See page 0]: <axis_config_1>
      - x:
        - column_id [See page 0]: <column_id_x_axis>
      - y:
        - column_id [See page 0]: <column_id_y_axis>

```

```
color [See page 0]:  
  - column_id [See page 0]: <column_id_color>  
axis_configs [See page 0]: <axis_config_2>  
locked [See page 0]: [ true | false ]  
client_state [See page 0]: <client_state_string>  
display_mode [See page 0]: <display_mode_string>  
guid [See page 0]: <answer_guid>
```

Syntax of the Pinboard TSL file

The `TSL` file for Scriptable Pinboards has a specific syntax.

See the [Parameters \[See page 547\]](#) section for details about the keywords used in this example.

You may not see each of these parameters in your own TSL files, depending on whether each variable is explicitly defined. For example, if you do not have any filters on your Pinboard, the `filters` parameter does not appear. You can add that variable to the TSL file to specify filters for your Pinboard.

```

pinboard [See page 0]:
  name [See page 0]: <pinboard_name>
  description [See page 0]:
    This is a multi-line description of the pinboard
    Description line 2
  visualizations [See page 0]:
    - answer [See page 0]:
      This section includes all the Answer specification for a visualization, from name to display_mode, in the Answer syntax \[See page 543\] section above.
      id [See page 0]: <viz_id_1>
    - answer [See page 0]:
      This section includes all the Answer specification for a second visualization. In this case, the visualization is a headline.
      id [See page 0]: <viz_id_2>
      display_headline_column [See page 0]: <headline_column>
    filters [See page 0]:
      - column [See page 0]: <filtered_column_name_1>
        oper [See page 0]: <filter_operator>
        values [See page 0]: <filtered_values>
          - value 1
          - value 2
      - column [See page 0]: <filtered_column_name_2>
    layout [See page 0]:
      tiles:
        - visualization_id [See page 0]: <visualization_id_1>
          size [See page 0]: <viz_id_1_size>
        - visualization_id [See page 0]: <visualization_id_2>
  guid [See page 0]: <pinboard_guid>

```

Parameters of TSL files

aggregation

The default aggregation of the Worksheet, View, or Table column, or the aggregation of the output for a formula.

Aggregation options depend on the data type.

Possible values: SUM , COUNT , AVERAGE , MAX , MIN , COUNT_DISTINCT , NONE , STD_DEVIATION , and VARIANCE

Default: SUM

alias

An alternate name for the table

answer

Top-level container for all object definitions within an Answer.

answer_columns

A list of columns generated by the search query.

axis_configs

Specifies the columns for each axis on a chart. If you are displaying a column chart with a line chart overlaying it, for example, you would need to specify more than one `axis_config`.

calendar

Specifies the calendar used by a date column

Can be the Gregorian calendar (`default`), a fiscal calendar, or any custom calendar.

See [Set up a custom calendar \[See page 94\]](#)

chart

Contains configuration for the Answer, if it displays in chart format.

chart_columns

A list of columns in the chart.

client_state

A JSON string with more advanced chart and table configuration.

color

Color to use for conditional formatting or for the columns of an Answer in chart form, in the form of a HEX value.

column

The id of the column being filtered on.

columns

The columns in the table.

column_id

The `id` of the Worksheet or View column.

For Answers, `column_id` refers to how the column appears in the query. For example, if you sorted by `Quarter` in your search, from the `Commit Date` column, the `column_id` of the column is `Quarter(Commit Date)`.

column_type

The type of data the column represents. For a formula, the `column_type` refers to the output of the formula.

Possible values: `MEASURE` or `ATTRIBUTE`

For Worksheets, the default is: `MEASURE`

For formulas, the default depends on the [data_type \[See page 0\]](#). If the data type is `INT` or `BIGINT`, the formula output's `column_type` defaults to `Measure`. If the data type is `BOOL`, `VARCHAR`, `DOUBLE`, `FLOAT`, `DATE`, `DATETIME`, or `TIME`, the formula output's `column_type` defaults to `Attribute`.

conditional_formatting

Conditional formatting for the chart or table of an Answer.

connection

A way to identify the external data warehouse connection that the table resides in.

currency_type

The source of currency type

One of:

- `is_browser` : true infer the currency data from the locale of your browser
- `column` : <column_name> extracts the currency information from a specified column
- `iso_code` : <valid_ISO_code> applies currency based on the ISO code; see ISO 4217
[Currency Codes \(https://www.iso.org/iso-4217-currency-codes.html\)](https://www.iso.org/iso-4217-currency-codes.html)

See [Set currency type \[See page 442\]](#)

custom_name

Optional display name for a column.

data_type

The data type of the formula output or column. If the data type is `INT` or `BIGINT`, the formula output's `column_type` defaults to `Measure`. If the data type is `BOOL`, `VARCHAR`, `DOUBLE`, `FLOAT`, `DATE`, `DATETIME`, or `TIME`, the formula output's `column_type` defaults to `Attribute`. The possible data types are `Boolean`, `Text`, `Date`, `Datetime`, `Time`, `Numeric`, and `Decimal`.

db

The database that a table resides in. Note that this is not the same as the data warehouse (Falcon, Amazon Redshift, or Snowflake, for example).

db_column_name

The name of the column in the database. Note that this database is not the same as the data warehouse (Falcon, Amazon Redshift, or Snowflake, for example).

db_table

The name of the table in the database. Note that this database is not the same as the data warehouse (Falcon, Amazon Redshift, or Snowflake, for example).

description

The text that describes an object: a `worksheet`, a `worksheet_column`, `answer`, `pinboard`, `view`, `view_column` and so on.

destination

Name of destination table or view of the join

display_mode

Determines whether the Answer displays as a chart or a table. Specify either `CHART_MODE` or `TABLE_MODE`.

display_headline_column

If the visualization is a headline, this parameter specifies the column the headline comes from.

expr

The definition of the formula

filters

Contains specifications for Pinboard, View, and Worksheet filters.

format_pattern

The format pattern string that controls the display of a number, date, or currency column

See [Set number, date, and currency formats \[See page 438\]](#)

formulas

The list of formulas in the Worksheet, View, or Answer.

Each formula is identified by `name`, the `expr` (expression), and an optional `id` attribute.

fqn

A GUID for the table name

geo_config

Specifies the geographic information of a column

One of:

- `latitude` : `true` for columns that specify the latitude
- `longitude` : `true` for columns that specify the longitude
- `country` : `true` for columns that specify the country
- `region_name` for specifying a region in a country

Uses two paired parameters:

- `country: <country_name>`
- `region_name: <region_name_in_UI>`, which can be State, Postal Code, District, and so on.

See [Add a geographical data setting \[See page 434\]](#)

guid

The GUID for the Answer, Pinboard, Table, Worksheet, or View. You can find this string of letters and numbers at the end of the URL for an object.

headline_aggregation

Specifies the type of headline aggregation. Can be `COUNT`, `COUNT_DISTINCT`, `SUM`, `MIN`, `MAX`, `AVERAGE`, or `TABLE_AGGR`.

id

Specifies the id of an object, such as `table_paths`, `formula`.

For Answers, `id` refers to how the column appears in the query. For example, if you sorted by `Quarter` in your search, from the Commit Date column, the `id` of the column is `Quarter(Commit Date)`. Refer to [Components of a Search Query \[See page 0\]](#) to understand syntax.

For formulas within Answers, `id` refers to the display name of the formula. If you do not give your formula a name, it appears as 'Untitled Formula'.

identity

Specifies the identity of a table, based on its `name`, `id`, and `fqn`.

index_priority

A value (1-10) that determines where to rank a column's name and values in the search suggestions
ThoughtSpot prioritizes columns with higher values.
See [Change a column's suggestion priority \[See page 432\]](#).

index_type

The indexing option of the Worksheet, View, or table column
Possible values: `DONT_INDEX` , `DEFAULT` (see [Understand the default indexing behavior \[See page 429\]](#)), `PREFIX_ONLY` , `PREFIX_AND_SUBSTRING` , and `PREFIX_AND_WORD_SUBSTRING`
Default: `DEFAULT`
See [Index Type Values \[See page 430\]](#)

is_additive

Controls extended aggregate options for attribute columns
For attribute columns that have a numeric data type (`FLOAT` , `DOUBLE` , or `INTEGER`) or a date data type (`DATE` , `DATETIME` , `TIMESTAMP` , or `TIME`)
Possible values: `true` or `false`
Default: `true`
See [Making an ATTRIBUTE column ADDITIVE \[See page 420\]](#)

is_attribution_dimension

Controls if the column is an attribution dimension
Used in managing chasm traps.
Possible values: `true` by default, `false` to designate a column as not producing meaningful attributions across a chasm trap
Default: `true`
See [Change the attribution dimension \[See page 444\]](#)

is_bypass_rls

Specifies if the worksheet supports bypass of Row-level security (RLS)
Possible values: `true` or `false`
Default: `false`
See [Privileges that allow users to set, or be exempt from, RLS \[See page 351\]](#)

is_hidden

The visibility of the column
Possible values: `true` to hide the column, `false` not to hide the column
Default: `false`
See [Hide a column \[See page 424\]](#)

is_one_to_one

Specifies the cardinality of the join
Possible values: `true` , `false`
Default: `false`

is_primary_key

Determines if the table column is the primary key. Can be `true` or `false` .

is_foreign_key

Determines if the table column is the foreign key. Can be `true` or `false` .

join

Specific join, used in defining higher-level objects, such as table paths

Defined as `name` within `joins` definition

join_path

Specification of a composite join as a list of distinct `join` attributes

These `join` attributes list relevant joins, previously defined in the `joins`, by name.

Default: `{}`

join_progressive

Specifies when to apply joins on a worksheet

Possible values: `true` when joins are applied only for tables whose columns are included in the search, and `false` for all possible joins

Default: `true`

See [How the worksheet join rule works \[See page 477\]](#)

joins

List of joins between tables and views, used by the Worksheet or View.

Each join is identified by `name`, and the additional attributes of `source`, `destination`, `type`, and `is_one_to_one`.

layout

Specifies the Pinboard layout, in the order that a `visualization_id` is listed.

locked

The 'automatically select my chart' option in the UI. If set to `true`, the chart type does not change, even when you add items to the query.

max

Maximum value for conditional formatting.

min

Minimum value for conditional formatting.

name

The name of an object. Applies to `worksheet`, `table`, `join`, `formula`, `answer`, `pinboard`, `view`, `table`, `connection` and so on.

For Answers, `name` refers to how the column appears in the query. For example, if you sorted by `Quarter` in your search, from the `Commit Date` column, the `name` of the column is `Quarter(Commit Date)`. Refer to [Components of a Search Query \[See page 0\]](#) to understand syntax.

on

The keys that your tables are joined on.

oper

The operator of the Pinboard, View or Worksheet filter. Accepted operators are `"in"`, `"not in"`, `"between"`, `<=`, `!=`, `<=`, `>=`, `>`, or `<`.

ordered_column_ids

A list of columns, in the order they appear in the table.

phrase

Phrase associated with a View column.

pinboard

Top-level container for all object definitions within the Pinboard.

properties

The list of properties of a Worksheet, Table, or View column, a Worksheet or View itself, or the properties of the output for a formula within an Answer, Worksheet, or View.

For Worksheets, Views, and Tables, each column can have the following properties, depending on its definition: `column_type`, `aggregation`, `index_type`, `is_hidden`, `index_priority`, `synonyms`, `is_attribution_dimension`, `is_additive`, `calendar`, `format_pattern`, `currency_type`, `geo_config`, `spotiq_preference`, and `search_iq_preferred`.

Worksheets and Views themselves can have the following properties that affect query generation: `is_bypass_rls`, and `join_progressive`.

For Answers, each formula's output can have the following properties, depending on its definition: `column_type` and `aggregation`.

plotAsBand

Specifies whether to plot the chart conditional formatting like a band on the Visualization. This is the 'fill chart' option in the UI.

query

The query that the View is based on.

range

Range for the conditional formatting to apply to, with a specified `min` and `max`.

rule

A conditional formatting rule.

schema

The schema that the table is a part of.

search_query

A string that represents the fully disambiguated search query. Refer to [Components of a Search Query \[See page 0\]](#) to understand syntax.

show_headline

Determines whether to show the headline for this column. `true` shows the headline.

size

The size of a visualization in a Pinboard. The options are `EXTRA_SMALL`, `SMALL`, `MEDIUM`, `LARGE`, `LARGE_SMALL`, `MEDIUM_SMALL`, and `EXTRA_LARGE`.

source

Name of source table or view of the join

spotiq_preference

Specifies whether to include a column in SpotIQ analysis. Specify `EXCLUDE`, or this property defaults to include the column in SpotIQ Analysis.

Refer to [Set columns to exclude from SpotIQ analyses \[See page 426\]](#).

synonyms

Alternate names for the column, used in search

See [Create synonyms for a column \[See page 424\]](#)

table

Top-level container for all object definitions within the Table.

Specific table, used in defining higher-level objects, such as table paths.

Defined as `name` within `tables` definition.

For Answers, this parameter contains configuration for the Answer, if it displays in table format.

table_columns

The columns in an Answer that is being displayed in table format.

table_paths

The list of table paths

Each table path is identified by the `id`, and additional attributes of `table` and `join_path`.

tables

List of tables used by the Worksheet or Answer.

Each table is identified by `name`.

type

For Worksheets and Views, this is the join type.

Possible values: `LEFT_OUTER` for left outer join, `RIGHT_OUTER` for right outer join, `INNER` for inner join, `OUTER` for full outer join

Default: `INNER`

For Tables, this is the Embrace connection type.

Possible values: `Snowflake` or `Amazon Redshift`.

For Answers, this is the chart type

Possible values: `COLUMN`, `BAR`, `LINE`, `PIE`, `SCATTER`, `BUBBLE`, `STACKED_COLUMN`, `AREA`, `PARETO`, `COLUMN`, `GEO_AREA`, `GEO_BUBBLE`, `GEO_HEATMAP`, `GEO_EARTH_BAR`, `GEO_EARTH_AREA`, `GEO_EARTH_GRAPH`, `GEO_EARTH_BUBBLE`, `GEO_EARTH_HEATMAP`, `WATERFALL`, `TREEMAP`, `HEATMAP`, `STACKED_AREA`, `LINE_COLUMN`, `FUNNEL`, `LINE_STACKED_COLUMN`, `PIVOT_TABLE`, `SANKEY`, `GRID_TABLE`, `SPIDER_WEB`, `WHISKER_SCATTER`, `STACKED_BAR`, or `CANDLESTICK`.

values

The values being filtered (excluded or included) in a Pinboard, View, or Worksheet.

view

Top-level container for all object definitions within the View.

view_columns

The list of columns in the View.

Each column is identified by `name`, `description`, `column_id`, `phrase` and `properties`.

visualizations

The visualizations in a Pinboard: tables, charts, and headlines.

visualization_id

The id of a visualization. Used to specify the Pinboard's [layout \[See page 0\]](#).

worksheet

Top-level container for all object definitions within the worksheet

worksheet_columns

The list of columns in the worksheet

Each worksheet is identified by `name`, `description`, `column_id`, and `properties`.

Limitations of working with TSL files

There are certain limitations to the changes you can apply by editing a Worksheet, Answer, table, View, or Pinboard through TSL.

- Formulas and columns can either have a new name, or a new expression. You cannot change both, unless migrating or updating the worksheet two times.
- It is not possible to reverse the join direction in the TSL script.
- You cannot create new tables using Scriptability. You can only update existing tables.
- You can only change logical tables using Scriptability. You cannot change the physical version of the table that exists in a database. When you change the `column_name`, for example, the name changes in the application, but not in the physical table in the database.
- You cannot import manually compressed .zip files. You can only import .zip files that you exported from ThoughtSpot: either an object and its associated data sources, or multiple objects of the same type that you exported from the object list page.

SpotApps

Summary: SpotApps, ThoughtSpot's scriptable applications, allow you to migrate multiple objects to and from clusters.

SpotApps take advantage of [Scriptability \[See page 520\]](#), ThoughtSpot's solution for exporting, enhancing, and migrating ThoughtSpot objects, to provide you with scriptable applications you can use in multiple clusters.

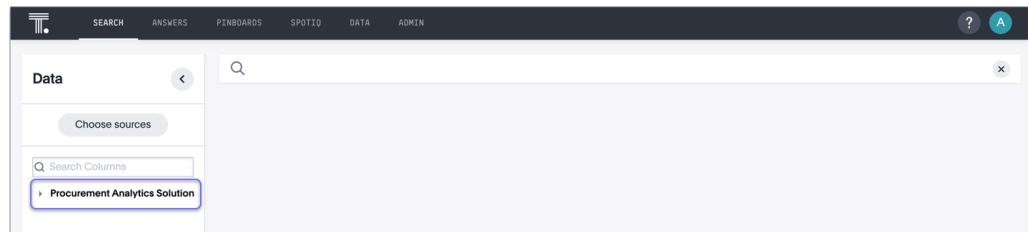
Once you connect to your data, you can work with your ThoughtSpot contacts to deploy ThoughtSpot's scriptable applications, or SpotApps, which provide an easy way for you to start getting value from your data.

ThoughtSpot offers [pre-built SpotApps \[See page 556\]](#), which leverage your data in Snowflake or Redshift to provide pre-built Pinboards, Answers, Views, Tables and Worksheets. You can also [create your own SpotApps \[See page 557\]](#) by exporting multiple objects of the same type in one zip file, or an object and its associated data sources, and migrating these objects to another environment.

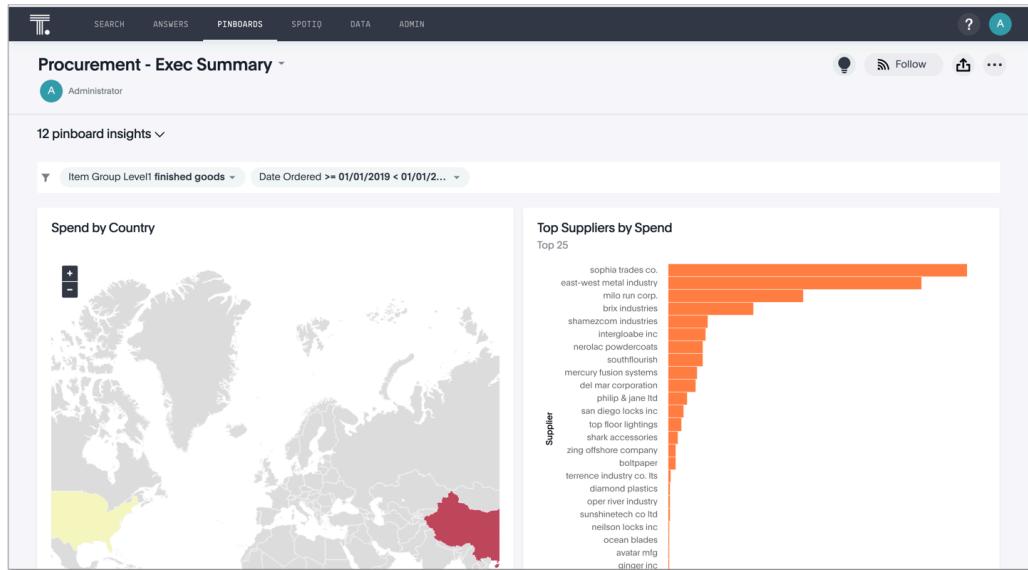
Pre-built SpotApps

ThoughtSpot offers 3 pre-built SpotApps: Salesforce, procurement, and accounts receivable. These applications leverage your data in Snowflake or Redshift to provide pre-built Pinboards, Answers, Views, Tables and Worksheets.

For example, if you choose to use the Procurement SpotApp, the **Search** page contains a Worksheet for your users to query on, called "Procurement Analytics Solution."



Your users may want to understand what Answers and Pinboards are before they start searching, or they may only want to view pre-built objects. They can view any of the pre-existing objects, such as this Exec Summary Pinboard:

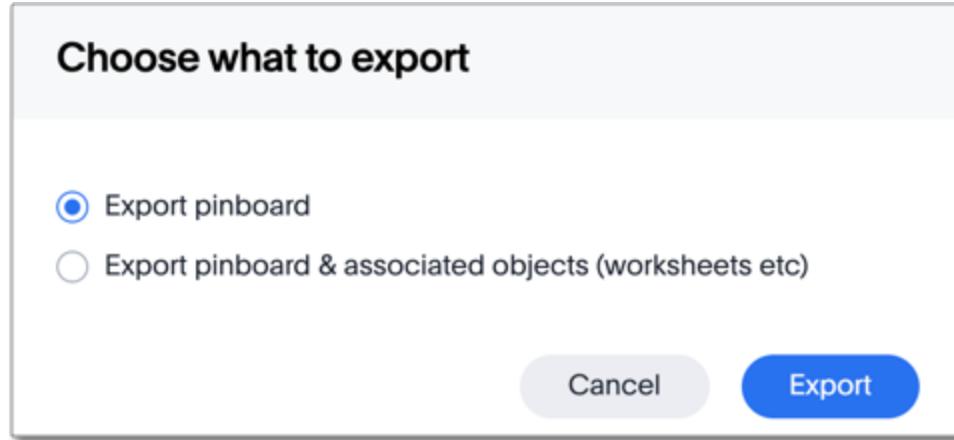


When you are ready to move to a production environment, you can migrate these Pinboards, Answers, Views, and Worksheets to your new environment using [Scriptability \[See page 520\]](#), ThoughtSpot's flat-file editing and migration system for ThoughtSpot objects.

Create and export SpotApps

You can create your own SpotApps in two ways:

You may have a specific Pinboard or other object that you would like to migrate to another cluster, but that cluster may not contain the Worksheets, Tables, or Views that the Pinboard's data comes from. When you export any Pinboard, Answer, Worksheet, or View, you have the option to export its associated data sources as well. This creates a SpotApp that you can migrate to another cluster without worrying about missing dependencies. The SpotApp .zip file contains a document called the `Manifest` file, which defines the objects you exported, and their underlying data sources.



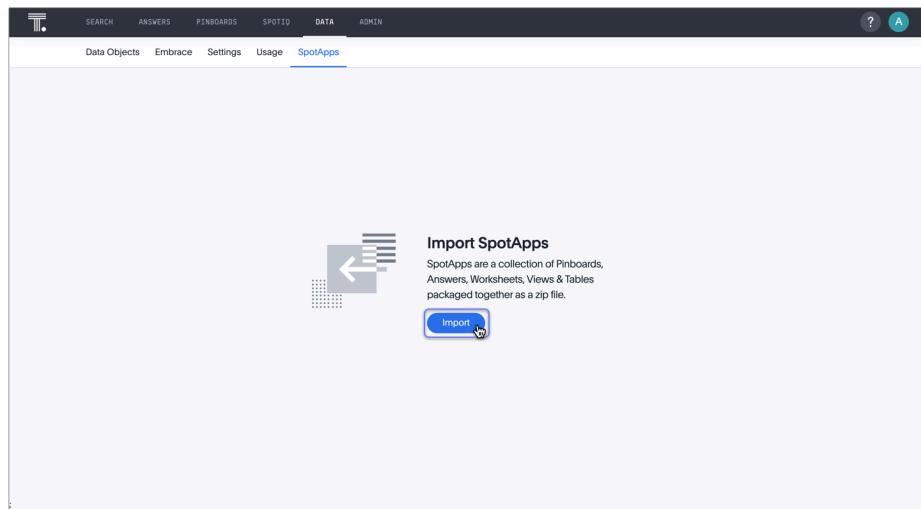
Similarly, you might want to migrate all the Answers on one cluster to another. You can select them all on the **Answers** list page, and export them as a SpotApp in .zip format. You can also choose to export their associated data sources in the same file.

See [Scriptability \[See page 520\]](#) for more information on exporting and importing objects.

Import SpotApps

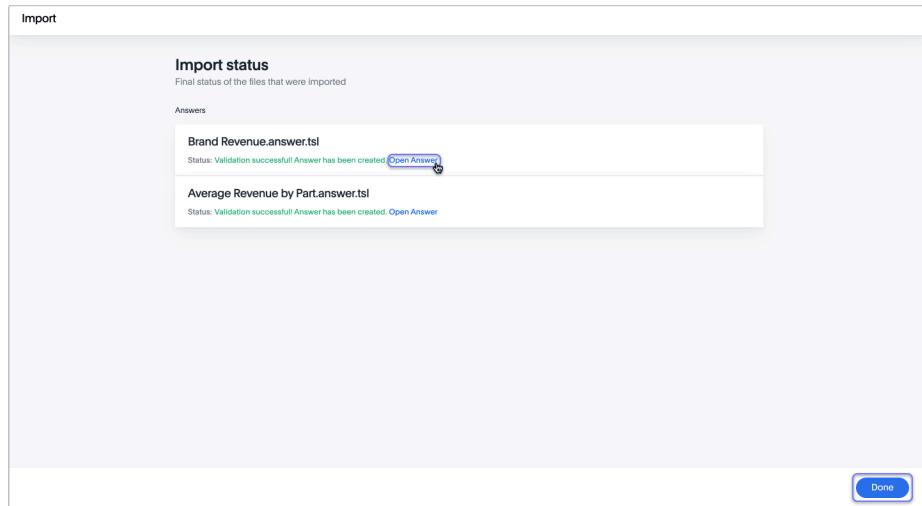
You can import SpotApps from the SpotApps page, under **Data > SpotApps**.

1. From the **SpotApps** page under the **Data** tab, click **Import**.



2. In the **Import** interface, click **Select .tsl or .zip files to upload**.

3. In your file system, find and select the .zip file for the SpotApp.
4. If you constructed the file correctly, the **Import** interface displays a *Validation successful* message. You can now import the file.
5. You can unselect any files in the `.zip` file you do not want to upload.
6. Click **Import selected files**.
7. The **Import Status** screen displays the status of the objects you imported. You can open the object(s) that you imported, or click **Done** to return to the main object page.



Limitations

You cannot import manually compressed .zip files. You can only import .zip files that you exported from ThoughtSpot: either an object and its associated data sources, or multiple objects of the same type that you exported from the object list page.

Enable SearchIQ

Summary: Enable SearchIQ to provide natural language search for some or all of your users.

Note: SearchIQ is in **Beta**

- SearchIQ is only available in English.
- The feature is off by default; to enable it, contact [ThoughtSpot Support](#) [See page 0].
- You must have the **Can use experimental features** permission.

SearchIQ is a different search experience that understands more natural, speech-like search language than the original ThoughtSpot search or [Search+](#) [See page 0]. For example, you can search for `What was my top selling product last month?`, instead of having to type `top 1 product by sales last month`. You can also speak your search using the voice-to-text capability of your operating system.

Users can easily [teach SearchIQ their language](#) [See page 0] as they use it. Over time, SearchIQ gets smarter about how to interpret terms that don't come directly from keywords, column names, or data values. In order to give it time to learn, the best practice recommendation is to enable SearchIQ for a small group of users initially. Over time you can roll it out to larger groups as SearchIQ learns what your users search for the most.

To enable SearchIQ on a ThoughtSpot instance:

1. Call ThoughtSpot Support and ask them to turn on SearchIQ for your ThoughtSpot instance.
2. Grant the **Can use experimental features** privilege to any group that should have access to the SearchIQ search experience.

Edit group

Group name *	Retail West
Display name *	Retail West
Sharing visibility *	SHARABLE
Description	
Privileges	<input type="checkbox"/> Can administer ThoughtSpot <input checked="" type="checkbox"/> Can upload user data <input checked="" type="checkbox"/> Can download data <input type="checkbox"/> Can share with all users <input checked="" type="checkbox"/> Can manage data <input checked="" type="checkbox"/> Can use experimental features <input type="checkbox"/> Can invoke Custom R Analysis <input checked="" type="checkbox"/> Has Spot IQ privilege <input type="checkbox"/> Can administer and bypass RLS

* Required field

Manage Groups Manage Users

No Groups in Group

Search by name

Clear all Select all

- Administrator
- Marketing
- Retail East
- SpotIQ
- ThoughtSPORT
- basic

Cancel UPDATE

Note: Because the SearchIQ search experience is very different from the original ThoughtSpot search experience, you should give your new users some orientation on it, so they know what to expect.

Related information

- [About SearchIQ \[See page 0\]](#)
- [Use SearchIQ \[See page 0\]](#)
- [Teach SearchIQ your language \[See page 0\]](#)

Optimize SearchIQ

Summary: For SearchIQ to work well, you must enable, optimize, and enable it on the data source.

SearchIQ is in **Beta**.

Note: SearchIQ is turned off by default, you can have ThoughtSpot Support enable it for you.

SearchIQ is only available in English.

SearchIQ is a search experience that understands a more natural, speech-like search language. Contrast it with the [original ThoughtSpot search \[See page 0\]](#). For example, you can search for *What was my top selling product last month?*, instead of typing *top 1 product by sales last month*. You can also speak your search using the voice-to-text capability of your operating system.

Data source owners must enable SearchIQ for their data sources, and then optimize the data source so users get better results. These are the primary steps for enabling and optimizing SearchIQ:

1. [Enable columns for SearchIQ \[See page 564\]](#)
2. [Add Experts for SearchIQ \[See page 570\]](#)
3. [Train SearchIQ \[See page 573\]](#)

Prerequisites

- SearchIQ is turned off by default; ask your administrator to enable it.
- You must be a member of a group that has the **Can use experimental features** permission.

Enable Columns

See [Enable Columns for SearchIQ \[See page 564\]](#).

Add Experts

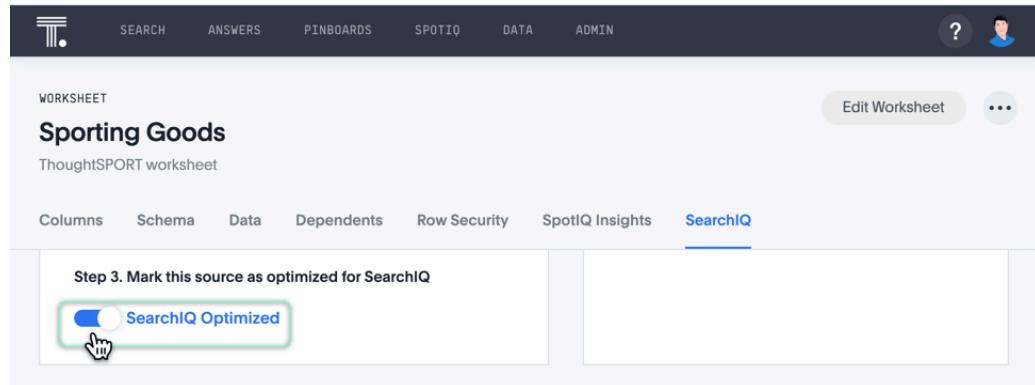
See [Add Experts for SearchIQ \[See page 570\]](#).

Train SearchIQ

See Train SearchIQ [See page 573].

Mark the source as optimized for SearchIQ

After you finish training SearchIQ on your data source, click the **Data** tab, navigate to the data source, select its **SearchIQ** tab, scroll to the bottom left corner, and finally turn on the setting for **SearchIQ Optimized**.



How does SearchIQ work for end Users

After you finish training SearchIQ on the Dataset, we recommend that you share with the users a list of the questions you used for training SearchIQ, the names of columns and their descriptions, common synonyms, and some guidance on how to ask a question.

Remember that users who experience SearchIQ see only the columns that SearchIQ knows.

Related information

- [About SearchIQ \[See page 0\]](#)
- [Use SearchIQ \[See page 0\]](#)
- [Teach SearchIQ your language \[See page 0\]](#)

Enable Columns for SearchIQ

Summary: Allow users to use SearchIQ by enabling SearchIQ for the columns of the data source."

SearchIQ is in **Beta**.

Note: SearchIQ is turned off by default. ThoughtSpot Support can enable it for you. SearchIQ is only available in English.

Before enabling columns, keep in mind these points, and other in-product guidelines:

- Enable no more than fifteen (15) columns for SearchIQ.
- Enable at most one (1) date column.
- Column names should not contain more than 3 words.
- Name the columns well, by optimizing for business users. For example, use *Opportunity Cost* and avoid `opp_cst`.
- Avoid indexing descriptive columns, ones that contain more than seven (7) words, or contain values with stop words, such as *is*, *are*, *the*, *what*, and so on). See the [Stop Words \[See page 0\]](#) reference for the full list of stop words.
- For columns that represent formulas with categorical values, set the **Indexing Type** to `PREFIX_AND_SUBSTRING`.

To enable columns in a worksheet for SearchIQ, follow these steps:

1. Click **Data** on top navigation bar.
2. Click **Tables**.
3. Click **Worksheets**.
4. Select and open the worksheet you want to enable for SearchIQ.
5. Scroll to the far right, until you see the **SearchIQ Enabled** attribute.

Set it to Yes for each column enabled for SearchIQ.

The screenshot shows the ThoughtSpot Admin interface for a 'Sporting Goods' worksheet. The 'Columns' tab is selected. A green box highlights the 'SEARCHIQ ENABLED' column, which contains checkboxes for each column. Most checkboxes are checked ('YES'), while the 'Department' column's checkbox is unchecked ('NO').

COLUMN NAME	ON	CALENDAR TYPE	ENTITY CATEGORY	SEARCHIQ ENABLED
Sales	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Gross Margin	NONE	DEFAULT	<input type="radio"/> NO	
Quantity	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
POS Transaction Nu...	NONE	DEFAULT	<input type="radio"/> NO	
Date	Retail-Calendar	DEFAULT	<input checked="" type="checkbox"/> YES	
Latitude	NONE	DEFAULT	<input type="radio"/> NO	
Longitude	NONE	DEFAULT	<input type="radio"/> NO	
Store City	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Store County	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Store Name	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Store State	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Store Zip Code	NONE	DEFAULT	<input checked="" type="checkbox"/> YES	
Department	NONE	DEFAULT	<input type="radio"/> NO	

6. Change the **Entity Category** for the columns you marked **SearchIQ Enabled**, from *default* to one that accurately describes the column.

See Set Entity Categories [See page 577]. The category options are *default*, *person*, *replace*, *time*, *money*, *product*, *zip_code*, *lat_long*, *company_org*, and *num_types*.

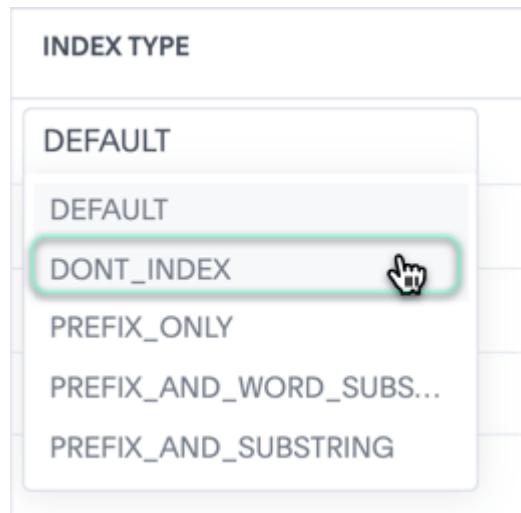
The screenshot shows the ThoughtSpot Worksheet interface. At the top, there are navigation links: SEARCH, ANSWERS, PINBOARDS, SPOTIQ, DATA, and ADMIN. On the far right, there are icons for Help, User Profile, and More. Below the navigation, it says "WORKSHEET" and "Sporting Goods". Underneath that, it says "ThoughtSPORT worksheet". There are several tabs: Columns (which is selected and highlighted in blue), Schema, Data, Dependents, Row Security, SpotIQ Insights, and SearchIQ. The main area is a table with the following columns: COLUMN NAME, ON, CALENDAR TYPE, ENTITY CATEGORY, and SEARCHIQ ENABLED. The rows contain the following data:

COLUMN NAME	ON	CALENDAR TYPE	ENTITY CATEGORY	SEARCHIQ ENABLED
Sales	NONE	MONEY	<input checked="" type="radio"/> YES	
Gross Margin	NONE	DEFAULT	<input type="radio"/> NO	
Quantity	NONE	NUM_TYPES	<input checked="" type="radio"/> YES	
POS Transaction Nu...	NONE	DEFAULT	<input type="radio"/> NO	
Date	Retail-Calendar	TIME	<input checked="" type="radio"/> YES	
Latitude	NONE	DEFAULT	<input type="radio"/> NO	
Longitude	NONE	DEFAULT	<input type="radio"/> NO	
Store City	NONE	PLACE	<input checked="" type="radio"/> YES	
Store County	NONE	PLACE	<input checked="" type="radio"/> YES	
Store Name	NONE	COMPANY_ORG	<input checked="" type="radio"/> YES	
Store State	NONE	PLACE	<input checked="" type="radio"/> YES	
Store Zip Code	NONE	ZIP_CODE	<input checked="" type="radio"/> YES	
Department	NONE	DEFAULT	<input type="radio"/> NO	

7. Add **Synonyms** for all columns you marked **SearchIQ Enabled**, simply by typing them in the appropriate cell. See [Add common synonyms \[See page 424\]](#) for more information.

COLUMN NAME	ADDITIVE	AGGREGATION	HIDDEN	SYNONYMS	SPOTIQ PREFERENCE
Sales	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	purchases, sale...	DEFAULT
Gross Margin	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Click to edit	DEFAULT
Quantity	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	amount, volume	DEFAULT
POS Transaction Nu...	<input checked="" type="radio"/> YES	SUM	<input type="radio"/> NO	Click to edit	DEFAULT
Date	<input type="radio"/> NO	NONE	<input type="radio"/> NO	when, year	DEFAULT
Latitude	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT
Longitude	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT
Store City	<input type="radio"/> NO	NONE	<input type="radio"/> NO	location, where	DEFAULT
Store County	<input type="radio"/> NO	NONE	<input type="radio"/> NO	county, where	DEFAULT
Store Name	<input type="radio"/> NO	NONE	<input type="radio"/> NO	name of store	DEFAULT
Store State	<input type="radio"/> NO	NONE	<input type="radio"/> NO	state, where	DEFAULT
Store Zip Code	<input type="radio"/> NO	NONE	<input type="radio"/> NO	zipcode, zip, wh...	DEFAULT
Department	<input type="radio"/> NO	NONE	<input type="radio"/> NO	Click to edit	DEFAULT

8. Change the **Index Type** to `DONT_INDEX` for columns when they must be excluded from indexing. We recommend excluding a column when it is descriptive in nature, =contains a large amount of text, has [stop words](#) [See page 0], or for a variety of similar reason. See [Turn off indexing](#) [See page 428].



This prevents SearchIQ from scanning through large amounts of text to understand what user asked. Note that searches on text fields that use the `contains` keyword still work.

WORKSHEET

Sporting Goods

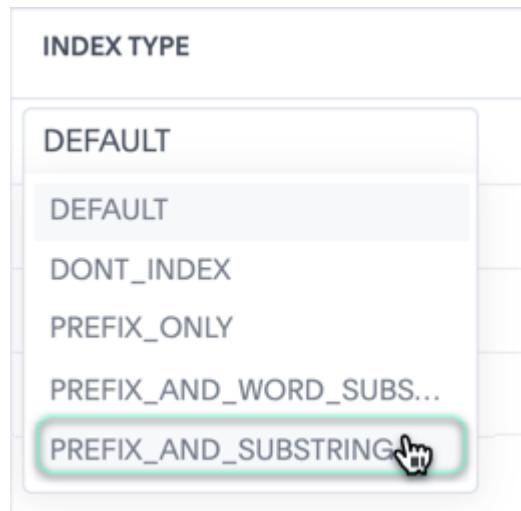
ThoughtSPORT worksheet

Save Changes Edit Worksheet ...

Columns Schema Data Dependents Row Security SpotIQ Insights SearchIQ

COLUMN NAME	INDEX TYPE	GEO CONFIG	INDEX PRIORITY
Share of Total Yearly Sales	DONT_INDEX	None	1
Sales Per Customer	DONT_INDEX	None	1
Sales	DEFAULT	None	1
Quantity	DEFAULT	None	1
Product Name	PREFIX_AND_SUBSTRING	None	1
POS Transaction Number	DEFAULT	None	1
Number of Children	DEFAULT	None	1
Net Margin	DEFAULT	None	1
Months Since Launch	DEFAULT	None	1
Margin Percentage	DONT_INDEX	None	1
Longitude	DEFAULT	Longitude	1
Latitude	DEFAULT	Latitude	1
Income Bracket	PREFIX_AND_SUBSTRING	None	1

- Change the **Index Type** for columns that represent formulas with categorical values to **PREFIX_AND_SUBSTRING**. See [Turn off indexing \[See page 428\]](#).



COLUMN NAME	INDEX TYPE	GEO CONFIG	INDEX PRIORITY
Share of Total Yearly Sales	DONT_INDEX	None	1
Sales Per Customer	DONT_INDEX	None	1
Sales	DEFAULT	None	1
Quantity	DEFAULT	None	1
Product Name	PREFIX_AND_SUBSTRING	None	1
POS Transaction Number	DEFAULT	None	1
Number of Children	DEFAULT	None	1
Net Margin	DEFAULT	None	1
Months Since Launch	DEFAULT	None	1
Margin Percentage	DONT_INDEX	None	1
Longitude	DEFAULT	Longitude	1
Latitude	DEFAULT	Latitude	1
Income Bracket	PREFIX_AND_SUBSTRING	None	1

10. Rename the columns to have names that are shorter (maximum three words), and easier for business users to understand.



11. Click **Save Changes**.

Add Experts for SearchIQ

Summary: It is useful to identify experts who can help with SearchIQ.

SearchIQ is in **Beta**.

Note: SearchIQ is turned off by default. ThoughtSpot Support can enable it for you. SearchIQ is only available in English.

We recommend that you add Experts to the datasource. These people in your organization know a great deal about the data, and can help you train SearchIQ. They are also the individuals who can assist business users in validating the answers they receive as a result of both regular Search and SearchIQ.

You can learn more about that by reading about our [Ask an expert \[See page 0\]](#) feature.

To specify the experts for the datasource, follow these steps:

1. Click the ellipsis icon at the top right corner, and select **Manage Experts**.

The screenshot shows a ThoughtSpot worksheet titled "Sporting Goods". At the top right, there is a modal window titled "Edit Worksheet" with a list of options: "Review Suggestions (2)", "SpotIQ analyze", "Manage Experts" (which has a hand cursor icon over it), and "Share". Below the modal is a table with columns: COLUMN NAME, ATTRIBUTION DIMENSION, CALENDAR TYPE, ENTITY CATEGORY, and SEARCHIQ ENABLED. The table contains ten rows of data, each with a "YES" button in the ATTRIBUTION DIMENSION column and a "NO" button in the SEARCHIQ ENABLED column.

COLUMN NAME	ATTRIBUTION DIMENSION	CALENDAR TYPE	ENTITY CATEGORY	SEARCHIQ ENABLED
Sales	<input checked="" type="button"/> YES	NONE	MONEY	<input checked="" type="button"/> YES
Gross Margin	<input checked="" type="button"/> YES	NONE	DEFAULT	<input type="button"/> NO
Quantity	<input checked="" type="button"/> YES	NONE	NUM_TYPES	<input checked="" type="button"/> YES
POS Transaction Nu...	<input checked="" type="button"/> YES	NONE	DEFAULT	<input type="button"/> NO
Date	<input checked="" type="button"/> YES	Retail-Calendar	TIME	<input checked="" type="button"/> YES
Latitude	<input checked="" type="button"/> YES	NONE	DEFAULT	<input type="button"/> NO
Longitude	<input checked="" type="button"/> YES	NONE	DEFAULT	<input type="button"/> NO
Store City	<input checked="" type="button"/> YES	NONE	PLACE	<input checked="" type="button"/> YES
Store County	<input checked="" type="button"/> YES	NONE	PLACE	<input checked="" type="button"/> YES
Store Name	<input checked="" type="button"/> YES	NONE	COMPANY_ORG	<input checked="" type="button"/> YES

2. In the **Manage Experts** window modal, you may see existing experts. Add more users and groups, and click **Add**.

Manage Experts

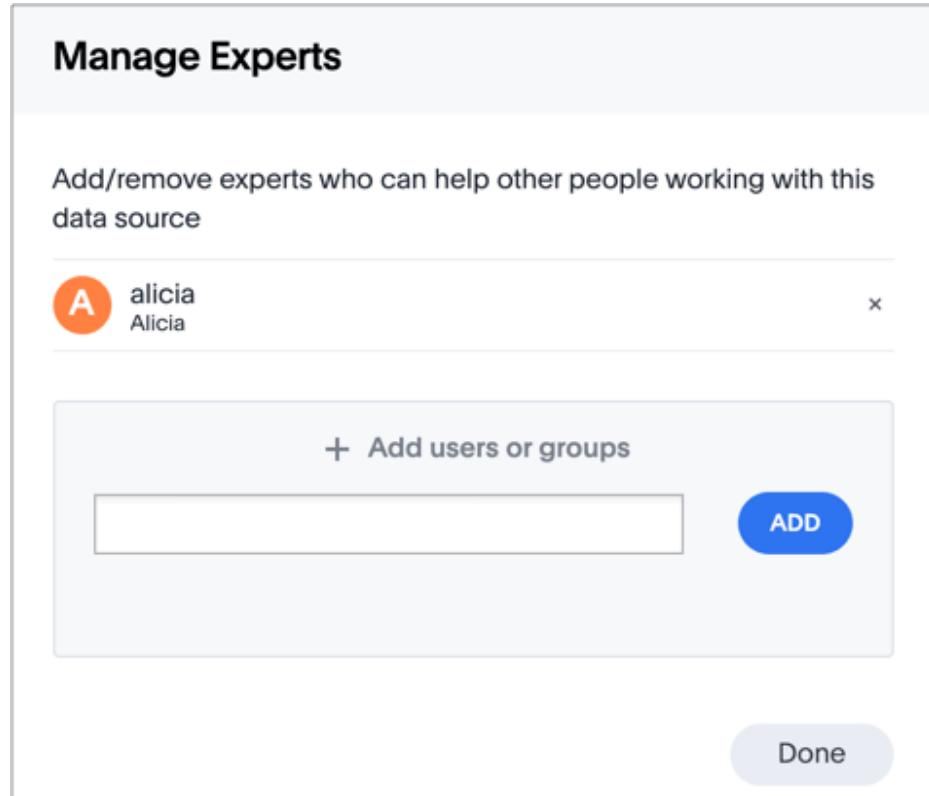
Add/remove experts who can help other people working with this data source

A alicia
Alicia x

+ Add users or groups

ADD

Done



3. After you finish adding experts, click **Done**.

Manage Experts

Add/remove experts who can help other people working with this data source

A alicia
Alicia

x

R Roza
Roza

x

S sean
Sean

x

+ Add users or groups

Done



Train SearchIQ

Summary: Training SearchIQ ensures a good natural language search experience.

SearchIQ is in **Beta**.

Note: SearchIQ is turned off by default. ThoughtSpot Support can enable it for you. SearchIQ is only available in English.

There are two complementary paths for optimizing SearchIQ to understand natural language search, [Data Modeling \[See page 573\]](#) and [Language Modeling \[See page 0\]](#). Also, consider the [Training Examples \[See page 574\]](#).

Data modeling settings

Add experts to the data source [\[See page 447\]](#) to enable the [Ask an Expert \[See page 0\]](#) workflow. This supplies users with another path to getting an answer when they get stuck.

Language modeling

We bundle SearchIQ with a set of pre-defined templates that help it understand natural language, and return the best possible answers. You can also add your own templates for each data source, and teach it new language interpretations:

Teach SearchIQ using a saved SearchIQ answer

See [Teach SearchIQ your language \[See page 0\]](#).

1. Select a saved answer.
2. Click one of the drop-down phrases that shows how SearchIQ interprets the search terms, and click **Teach**.

Use new searches to teach SearchIQ

We recommend that you train SearchIQ by performing at least 50 natural language searches. This teaches SearchIQ how to map common questions to relevant entities. Use points described in [Teach SearchIQ \[See page 0\]](#) for any unexpected answers.

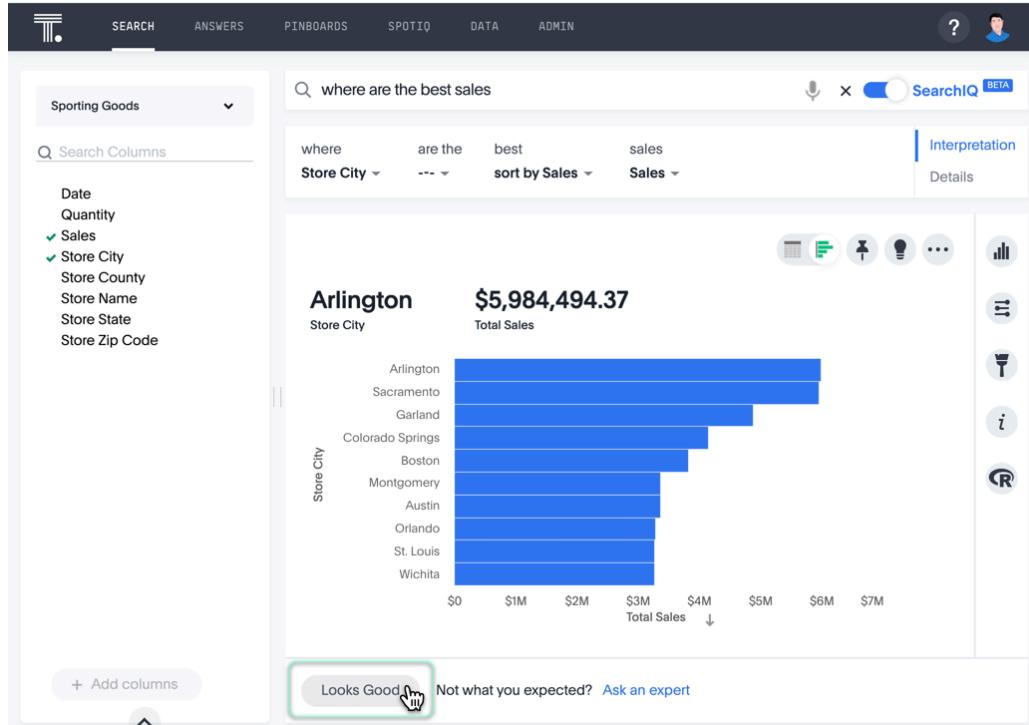
- For best results, use simple and short queries.
- Ask questions that contain only exact column names or their synonyms.
- Use exact values for filtering.
- Use keywords. See [Keywords \[See page 0\]](#).
- Use stopwords. See [Stop Words \[See page 0\]](#).

When SearchIQ returns a satisfactory table or chart, remember to click the **Looks Good** button to provide the necessary feedback.

Training Examples for SearchIQ

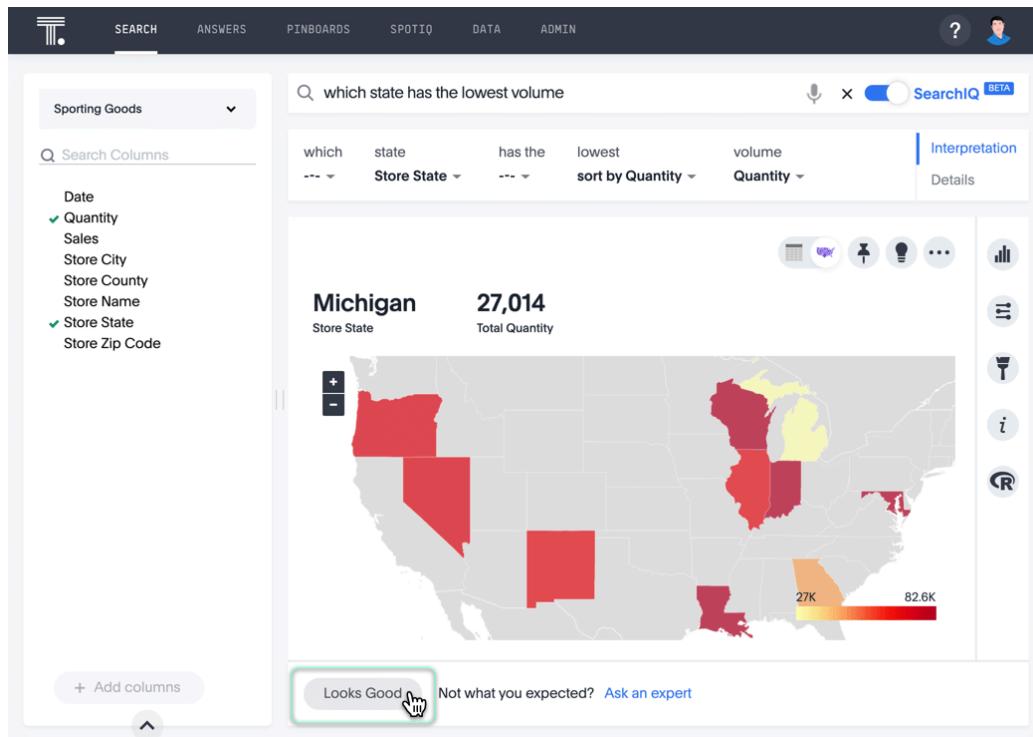
Example 1: where are the best sales

In this example, the query returned a very good answer, by interpreting *where* as *Store City* and *best* as *sort by Sales*. The chart **Looks Good**, and this is valuable feedback for training SearchIQ.



Example 2: which state has the lowest volume

In this example, the query returned a very good answer, by interpreting *lowest* as *sort by Quantity* and *volume* as *Quantity*. This chart also **Looks Good**, and provides valuable feedback for training SearchIQ.



Set entity categories for SearchIQ

Summary: Entity categories help describe the column content, and assist SearchIQ in interpreting natural language queries.

Note: SearchIQ is Beta

- SearchIQ is only available in English.
- SearchIQ is in Beta, and off by default; to enable it, contact ThoughtSpot Support [See page 0].
- You must have the **Can use experimental features** permission.

Entity categories specify how to categorize the data in the column, by entity type. By default, `ENTITY_TYPE` is not set. Entity categories support SearchIQ [See page 0] so that when you type a natural question, ThoughtSpot better knows how to interpret it. For example, if you ask “who are the top performers?” ThoughtSpot will first choose columns set with `PERSON` from which to return answers. If you ask “when was the movie Jurassic Park released?”, columns set to `TIME` will be used to answer the “when” part of the question, and so forth.

If you have access to tables and worksheets for data modeling purposes, you can specify entity categories for their columns. This lets you designate that column as representing a person, place, or thing. So then when a user asks “Which”, “Who”, or “Where”, an appropriate response can more easily be found in the data.

If you’re not sure how to change the data modeling settings, see the [Overview of data modeling settings](#) [See page 413].

About Entity Categories

These are the available Entity Categories:

Category	Description
PERSON	Contains data that represents a person, relevant to questions about “who?”

Category	Description
PLACE	Contains data that represents a location, relevant to questions about “where?”
TIME	Contains data that represents a date or time, relevant to questions about “when?”
PRODUCT	Contains data that represents a product
ZIP_CODE	Contains zip code data, relevant to questions like “where?” or “what zip code?”
LAT_LONG	Contains data that represents geographical positioning, relevant to questions like “where?”
COMPANY_ORG	Contains data that represents a company or organization
NUM_TYPES	Contains numerical data

Set Entity Categories

To specify entity categories:

1. Click **Data** in the top menu, and choose **Tables or Worksheets**.
2. Click the name of your table or worksheet.
3. On the **Columns** tab, find the COLUMN NAMES for which you want to specify entity categories, and scroll to the right to find **ENTITY CATEGORY**.
4. Use the drop-down menu to set the **ENTITY CATEGORY** to the type you want.
5. Click **SAVE CHANGES** in the upper right.

Related information

- [SearchIQ \[See page 0\]](#)
- [Overview of data modeling settings \[See page 413\]](#)

Edit the SearchIQ mappings

Summary: In SearchIQ, you can change the language mappings established earlier, during initial configuration.

Note: SearchIQ is in Beta

- SearchIQ is only available in English.
- SearchIQ is Beta, and off by default; to enable it, contact ThoughtSpot Support [See [page 0](#)].
- You must have the **Can use experimental features** permission.

When a user types a term in the SearchIQ search bar, SearchIQ interprets what the user means through language mappings. Users can teach their language to SearchIQ using [Teach SearchIQ your language](#) [See [page 0](#)].

As an administrator, you can see and change these language mappings by editing the mappings file. You can download the mappings file there from **Data > Settings > Language Mappings**. This allows you to view and edit all the language mappings users made, in bulk.

Overview of the mapping process

The language mappings in SearchIQ are controlled by a CSV (comma separated values) file called the mapping file. The mapping file is saved with an .xls file extension, so you can open and edit it using Excel. To make these changes you will download the mapping file, change the mappings, and upload your changes back into the system.

In each row of the mapping file, you will see a natural language search term, that is mapped to a search keyword, or something that occurs in the data (like a column name or value). You can use the mapping file to remove mappings that have been made in error. Remember these important guidelines when editing the Mappings file:

- Do not delete rows from the file. Instead, type “True” in the **Delete Y/N** column for any mapping you want to delete.
- Make sure to keep the file in the same format as it had when you downloaded it.

The model file contains a row for each natural language mapping a user has created using the Teach feature of SearchIQ. It isn't unusual to have tens of thousand of rows in this file.

Download the mapping file

Before you can make changes to the mapping file, you need to download it. Then, you edit it using Microsoft Excel, vi/vim, or a similar text editing tool.

To obtain the model file:

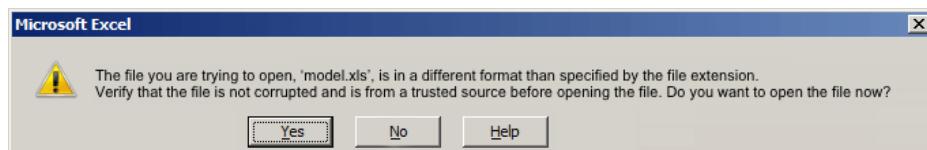
1. Log in to ThoughtSpot from a browser as an Administrator user.
2. Click **Data** in the top navigation bar.
3. Click **Settings**, then click **Language Mappings**.
4. Click **Download**.

Edit the file and change the settings

You can delete any of the mappings in the file. To delete a mapping:

1. Open the model file you downloaded (`mappings.xls`) in Excel, vi/vim, or a text editor.

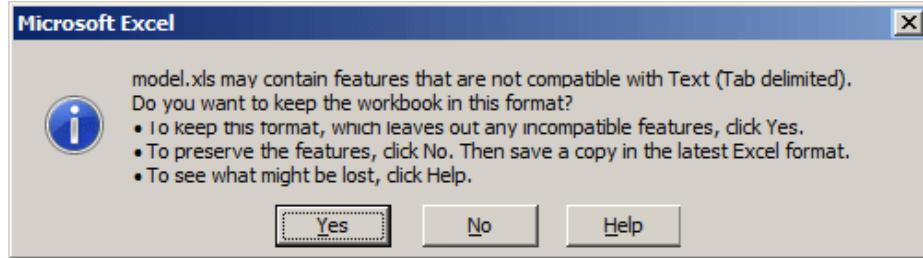
If you are using Excel, you may see a warning message similar to this.



Click **YES** to proceed.

2. Find the row for which you want to delete the mapping.
3. Find the **Delete Y/N** column.
4. Type **True** in the corresponding cell.
5. After making all your changes, save the model file.

If you are using Excel, you will see a message similar to this one. Click **YES** to save the file.



The mapping file must be saved as UTF-8 encoded. If your mapping file includes multi-byte characters, edit the file using vi or vim to ensure the file is saved in the correct format.

Otherwise, you won't be able to upload it after making your edits.

Upload the edited file

After you have made changes to the mapping file, you must upload it back to ThoughtSpot before the changes will take effect. To upload the mapping file:

1. Log in to ThoughtSpot from a browser as an Administrator user.
2. Click **Data**, on the top navigation bar.
3. Click **Settings**, then click **Language Mappings**.
4. Click **Browse your files** to upload the mappings.xls file, or drag and drop it in the zone.

If you receive an error message upon uploading the file, check that it does not include any multi-byte characters (for example, Japanese or other multi-byte language characters). If it does, you must save it as UTF-8 encoded. You may need to download the file again and make your edits using vi or vim.

As soon as the file is uploaded, your new settings will be reflected within a few minutes.

Related information

- [Teach SearchIQ your language \[See page 0\]](#)

Job management (scheduled pinboards)

Summary: All jobs on your cluster will appear on the Jobs Management page. You can also view jobs for individual pinboards under the pinboard Actions dropdown.

The **Jobs Management** page found in the **Data > Usage** section in the ThoughtSpot web application allows you to create and manage jobs, namely scheduled pinboards. Scheduled pinboards should help with preparing for recurrent meetings, when reviewing the same pinboard is necessary. They should also be useful when you have metrics you want to monitor at a consistent interval, like daily or monthly sales targets.

You can get pinboards emailed to you on a regular basis and do analysis offline. This introduces an additional format for you to consume and share pinboards with others, including those who don't have a ThoughtSpot account.

Run `tscli scheduled-pinboards enable` to enable scheduled pinboards on your cluster.

Scheduled pinboard creators

Administrators and users with `can schedule pinboard` privilege can schedule and manage pinboard jobs. These scheduled pinboard creators must have at least edit-only and view-only rights to the pinboard they want to share.

⚠ Warning: It is recommended that admins carefully choose who to give `can schedule pinboard` privilege to, since there is a possible security hole where a user with limited access can get a pinboard email with all access data.

Row level security

The scheduled pinboards respect row level security rules. This means if the recipients are users in ThoughtSpot, then they can only see data based on their own access to the pinboard. If the user does not have at least view-only access to the pinboard, then they will not see anything in the email. However, if the recipients are from outside of the cluster, then they will have access to the dataset of the pinboard based on the sender's permissions.

Scheduled pinboard formats

The pinboard visualizations are attached to the scheduled email as CSV or PDF files. Saved configurations such as pinboard filters are applied to the attachments. Refer to the table to see how the pinboard data is represented in each file format.

CSV	PDF
The CSV file gets data only for table visualizations.	The PDF file gets data for all visualizations.
The email has n CSV attachments, where there are n table visualizations in the pinboard.	The email has only one attachment file, which includes every visualization on its own page.
Table visualizations have all data rows that they're supposed to have.	Table visualizations include only the first 100 rows.
In the case of a corrupted pinboard: no email is sent. An error message indicating failure to export data is visible on the Data > Usage > Jobs Management page.	In the case of a corrupted pinboard: the PDF attachment has empty/error screenshots.
In the case of a corrupted visualization: an email with the visualizations whose data can be exported is sent. An error message indicating visualization export error is visible on the Jobs Management page.	In the case of a corrupted visualization: the PDF attachment has empty/error slots for the corrupted visualizations.

The size of each email is limited to 25 MB, which matches most email services size limitations.

And the total number of recipients for a scheduled pinboard job cannot exceed the default of 1000.

Related information

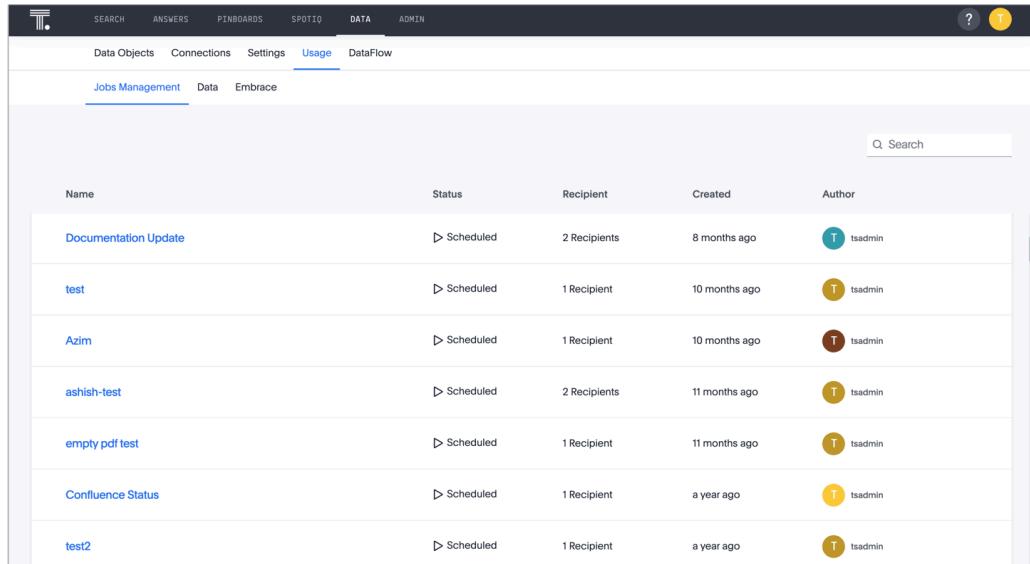
For information on creating a pinboard, see the [Schedule a pinboard job \[See page 0\]](#).

Scheduled pinboards management

Summary: You can manage all scheduled pinboards on the Jobs Management page.

Users who are not admins, but have the **can schedule pinboard** privilege, can only view pinboard schedules they've created. You can select specific jobs and choose to pause, resume, edit, or delete them. You can have up to 50 scheduled jobs on your cluster at time. Contact ThoughtSpot Support if you'd like to increase this limit.

View the Jobs Management page by clicking **Data > Usage > Jobs Management**.



The screenshot shows the ThoughtSpot interface with the navigation bar at the top. Under the 'DATA' tab, the 'Usage' sub-tab is selected. Below the navigation, there are three tabs: 'Jobs Management' (which is active), 'Data', and 'Embrace'. A search bar labeled 'Q. Search' is present. The main area displays a table of scheduled pinboards with the following columns: Name, Status, Recipient, Created, and Author. The table contains the following data:

Name	Status	Recipient	Created	Author
Documentation Update	> Scheduled	2 Recipients	8 months ago	tsadmin
test	> Scheduled	1 Recipient	10 months ago	tsadmin
Azim	> Scheduled	1 Recipient	10 months ago	tsadmin
ashish-test	> Scheduled	2 Recipients	11 months ago	tsadmin
empty pdf test	> Scheduled	1 Recipient	11 months ago	tsadmin
Confluence Status	> Scheduled	1 Recipient	a year ago	tsadmin
test2	> Scheduled	1 Recipient	a year ago	tsadmin

Bulk actions

Select the scheduled pinboards and use the **Delete**, **Resume**, and **Pause** buttons to perform these bulk actions. Deleting a pinboard will also delete any schedules linked to it.

Job statuses

Clicking on the row of a job will open a detailed view of every generated update of that job. You can see the start and end times of the job, as well as the status. Clicking on a job will show more information about the status updates.

The screenshot shows the ThoughtSpot interface with the following navigation bar:

- SEARCH
- ANSWERS
- PINBOARDS
- SPOTIQ
- DATA
- ADMIN

The DATA tab is selected, showing sub-options: Data Objects, Connections, Settings, Usage (which is highlighted), and DataFlow.

The sub-menu under Usage is "Jobs Management".

The main content area displays a table titled "Documentation Update" with the following data:

Name	Started at	Ended at	Status
test	17 hours ago	17 hours ago	Failed
Azim	2 days ago	2 days ago	Failed
ashish-test	-	-	-
empty pdf test	-	-	-
Confluence Status	-	-	-
test2	-	-	-

A search bar labeled "Q Search" is located in the top right corner of the content area.

Pinboard links

Click the scheduled pinboard name link to jump to an Edit schedule page, where you can edit the schedule configurations.

You can also click the pinboard link provided in the scheduled pinboard emails to jump to the pinboard in ThoughtSpot. In order to have the link direct you to the correct URL, you must first configure front end host and port access. Contact ThoughtSpot Support to configure these settings.

Introduction to monitoring

Summary: Learn how to monitor your system.

System monitoring tools in ThoughtSpot include several Pinboards in the **Admin Console** and system logs. Additionally, ThoughtSpot provides several worksheets and out-of-the-box system monitoring pinboards. From these worksheets, you can create your own custom visualizations and pinboards. This page introduces these features and directs you towards more detailed information.

Admin Console

The ThoughtSpot application includes an **Admin Console** center, where you can easily monitor usage, alerts, events and general cluster health. Navigate to the Admin Console by selecting **Admin** from the top navigation bar.

Only users with administrative privileges can view the **Admin Console**. However, administrative users can present the information in the **Admin Console** to others.

Administrators can also create their own, custom boards that reflect system data in ways that are meaningful to specific departments or groups. For more information, see the following documentation:

- [Health Overview board \[See page 589\]](#)
- [Data board \[See page 603\]](#)
- [Cluster Manager board \[See page 606\]](#)
- [Alerts and Events board \[See page 606\]](#)
- [System worksheets \[See page 611\]](#)
- [System pinboards \[See page 613\]](#)
- [Falcon monitoring Pinboards \[See page 616\]](#)
- [Performance Tracking Pinboard \[See page 621\]](#)

Much of the data presented by these boards is also available through `tscli` [commands \[See page 0\]](#).

Log files

Many of the administration commands output logging information to log files. The logs get written into the fixed directory `/export/logs`, with a sub-directory for each subsystem. The individual log directories are the following:

- `/export/logs/orion`
- `/export/logs/oreo`
- `/export/logs/hadoop`
- `/export/logs/zookeeper`

You can also view [additional topics that also touch on [log files \[See page 0\]](#)] throughout the documentation.

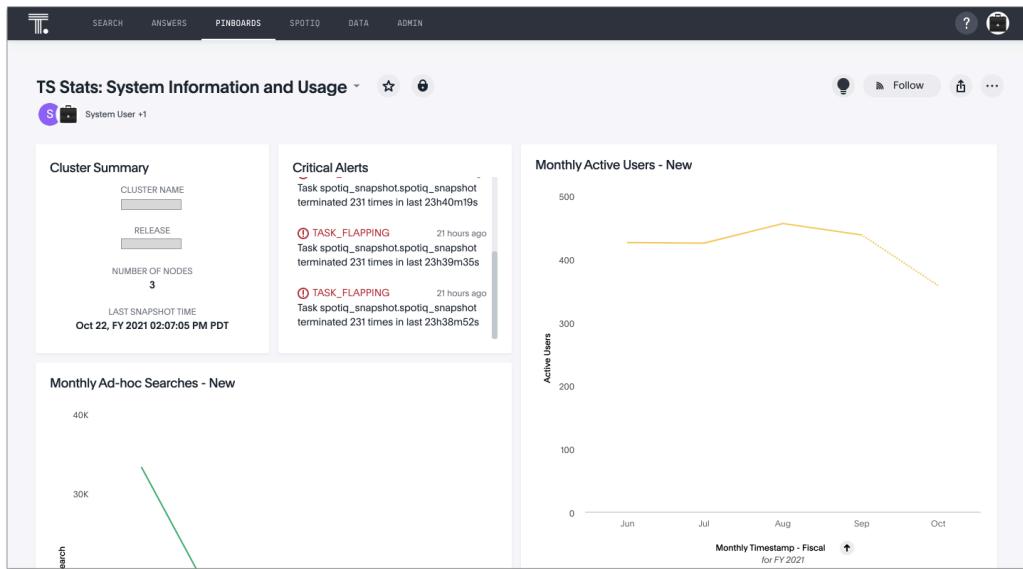
System monitoring notifications

You can configure ThoughtSpot to send emails to addresses you specify with monitoring reports and a cluster heartbeat. Follow these steps to [Set up monitoring \[See page 137\]](#).

System Information and Usage Pinboard

Summary: The System Information and Usage Pinboard provides useful information about your cluster.

You can view system and usage information for your ThoughtSpot cluster from the **TS Stats: System Information and Usage Pinboard**. This Pinboard contains several charts and tables that track monthly active users, top Pinboards by number of views, today's users, and so on. To view this Pinboard, search for it in the **Pinboards** tab.



Understand the TS Stats: System Information and Usage Pinboard

The **TS Stats: System Information and Usage Pinboard** page consists of standard ThoughtSpot Answers that update in real time and rely on internal system data. The Answers rely on underlying system worksheets, which are available to ThoughtSpot administrators. The information in these worksheets updates hourly from internal tables that collect monitoring statistics.

The **more options menu**  for the Answers in this Pinboard contains fewer options than the traditional menu. You can **SpotIQ analyst**, **download**, **edit**, or **present** the Answer. When you **edit** the Answer, you see the query that ThoughtSpot built the Answer on. You can also copy the Answer's embed link, use **Answer Explorer** [See page 0] on the Answer, or interact with the Answer by **drilling down** [See page 0]. While you can interact with and change a copied version of the Answer, you cannot save changes to the underlying query.

The following charts, tables, and panels are available on the **TS Stats: System Information and Usage Pinboard**: [cluster summary](#) [See page 590], [critical alerts](#) [See page 593], [monthly active users](#) [See page 596], [monthly ad-hoc searches](#) [See page 596], [monthly Pinboard views](#), [See page 597], [top users last month](#) [See page 598], [top Pinboards last month](#) [See page 599], [ad-hoc search by user](#) [See page 0], [today's users](#) [See page 600], and [configuration events](#) [See page 601].

Cluster Summary

This system panel contains basic information about your cluster. The **NUMBER OF NODES** is the number of installed nodes. This doesn't reflect the active nodes which may be more or less.

This summary includes the **LAST SNAPSHOT TIME** it reflects whether regular snapshots of your cluster are collected. This value should update regularly in real time. If you do not see it change or empty, you should check your cluster snapshot policy using the `tscli` command:

```
$ tscli snapshot-policy show
schedule {
    period {
        number: 1
        unit: HOUR
    }
    retention_policy {
        bucket {
            time {
                number: 1
                unit: HOUR
            }
            capacity: 3
        }
        bucket {
            time {
                number: 4
                unit: HOUR
            }
            capacity: 2
        }
    }
    offset_minutes_from_sunday_midnight: 0
}
enabled: false
```

You can see this policy is disabled, which is a problem. Production clusters should enable the default snapshot policy. When you show or enable the snapshot policy, you can see your `tscli` command reflected in the **Configuration Events** panel on this same page.

Display the features used in a cluster configuration

1. Log in to the ThoughtSpot cluster as the `admin` user.
2. Use the `tscli feature` subcommand to display your current configuration.

\$ tscli feature get-all-config	NAME	STATUS	CONFIGUR
ACTION			
Firewall		Disabled	
Saml		Disabled	
Ldap		Disabled	
CustomBranding		Disabled	
CustomBrandingFontCustomization		Disabled	
DataConnect		Disabled	
RLS		Enabled	
Callhome		Enabled	
SSHTunnel		Enabled	
Fileserver		Disabled	

Relational Data Cache

This section reports real-time information about tables in your cluster. Worksheet data is not included.

Value	Description
TABLES LOADED	Number of currently loaded tables.
TABLES BEING UPDATED	Number of table loads in-progress.
NEW TABLES BEING LOADED	Number of tables being loaded for the first time.

Value	Description
ROWS	Number of rows combined across all tables in ThoughtSpot.

Relational Search Engine

Value	Description
TABLES SEARCHABLE	Tables that are indexed and can be searched.
TABLES BEING INDEXED	Total of in-progress table indexing.
NEW TABLES BEING INDEXED	Total of first-time, in-progress table indexing.
TOKENS SEARCHABLE	Number of tokens of all table (combined) indexed in ThoughtSpot.

Critical Alerts

Displays critical and warning alerts. This includes when an alert was generated and from which service and machine. Administrators can get a custom report by issuing a `tscli alert list` on the appliance:

```
tscli alert list --since 4w
```

The critical alerts you can encounter in this display are the following:

- `TASK_FLAPPING`

```
Msg: Task {{.Service}}.{{.Task}} terminated {{._actual_num_occurrences}}
times in last {{._earliest_duration_str}}
```

This alert is raised when a task is crashing repeatedly. The service is evaluated across the whole cluster. So, if a service crashes 5 times in a day across all nodes in the cluster, this alert is generated.

- OREO_TERMINATED

```
Msg: Oreo terminated on machine {{.Machine}}
```

This alert is raised when the Oreo daemon on a machine terminates due to an error. This typically happens due to an error accessing Zookeeper, HDFS, or a hardware issue.

- HDFS_DISK_SPACE

```
Msg: HDFS has less than {{.Perc}}% space free
```

Raised when a HDFS cluster is low on total available disk space.

- ZK_INACCESSIBLE

```
Msg: Zookeeper is not accessible
```

Raised when Zookeeper is inaccessible.

- PERIODIC_BACKUP_FLAPPING

```
Msg: Periodic backup failed {{._actual_num_occurrences}} times in last  
{{._earliest_duration_str}}
```

This alert is raised when a periodic backup failed repeatedly.

- PERIODIC_SNAPSHOT_FLAPPING

```
Msg: Periodic snapshot failed {{._actual_num_occurrences}} times in last  
{{._earliest_duration_str}}
```

This alert is raised when periodic snapshot failed repeatedly.

- APPLICATION_INVALID_STATE_EXTERNAL

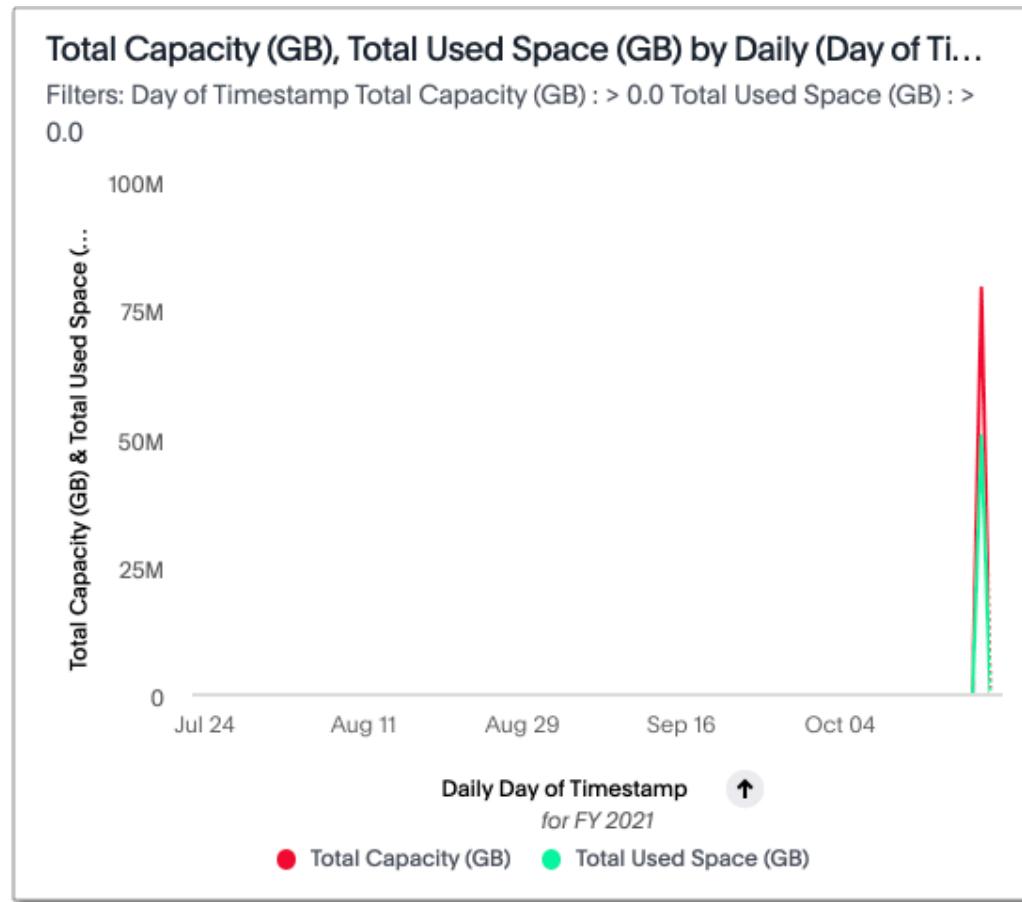
```
Msg: {{.Service}}.{{.Task}} on {{.Machine}} at location {{.Location}}
```

Raised when Application raises invalid state alert.

The possible alert types are CRITICAL, WARNING, ERROR, and INFO. For a full reference, see the Alert code reference [See page 0].

Space Utilization

The **Space Utilization** chart displays your cluster's total capacity and estimated used capacity over time.



The x-axis is by time and the y-axis measures the size in GB. You can zoom in and see daily or hourly utilization data. So, in the **Space Utilization** chart above, the green line shows the amount of capacity in use in the system, while the red line shows the total capacity. An increase in the red line at the end of a time period indicates the addition of extra hardware, resulting in increased capacity.

The query for this Answer is the following:

```
day of timestamp
total capacity (gb)
total used space (gb)
daily
last 90 days last 25 hours
total capacity (gb) > 0
total used space (gb) > 0
```

The chart relies on the `TS: Internal Table Wise Capacity WS` worksheet. It tracks total used space, which consists of raw uncompressed data, including replication.

Monthly Active Users

This chart shows the number of active users in the system over the last four months, and during the current month. An active user is defined as a user who has logged in at least one time in the defined time interval, in this case months.

The query for this Answer is the following:

```
monthly
last 4 months this month
active users
user != {null}
```

This query relies on the `TS: BI server` worksheet.

Monthly Ad-hoc Searches

Number of ad-hoc searches (queries) issued per month. An ad-hoc query is defined as any search or change to a search that builds a new Answer (result). An ad-hoc search can also be generated through SpotIQ or another UI/API interaction.

ThoughtSpot considers all of the following as ad-hoc searches (queries):

- User edits tokens (boxed terms) in the search bar.
- User opens an existing saved Answer and makes changes to tokens in the search bar.

- User opens an existing saved pinboard, edits a visualization, and makes change to the search tokens.
- Searches initiated by an API call for data with runtime filters

It is not considered a search (query) in this context if a user opens an existing saved aggregated worksheet and makes changes to its underlying query.

The query for this Answer is the following:

```
ad-hoc search
user action = 'answer_pinboard_context' 'answer_saved' 'answe
r_unsaved'
monthly
last 4 months this month
```

This Answer relies on data from the `TS: BI Server` worksheet.

Monthly Pinboard Views

Number of times a saved pinboard is viewed by a user. These scenarios are considered pinboard views:

- User opens an existing saved pinboard.
- User opens an embedded pinboard from a URL.
- Pinboard data is accessed using the an API.

These scenarios are not considered pinboard views:

- A user opens SpotIQ tab pinboards.
- A user opens admin tab pinboards.
- The system loads a pinboard on the homepage.
- The system loads the ‘learn how to use ThoughtSpot’ pinboard.

The query underlying this Answer is:

```
pinboard views
user action = 'pinboard_embed_view' 'pinboard_tspublic_no_runtime_filter' 'pinboard_tspublic_runtime_filter' 'pinboard_view'
monthly
last 4 months this month
```

The query uses the `TS: BI Server` data source.

Top Users Last Month

This Answer shows the top ThoughtSpot users ranked by number of actions the users performed in the last 30 days. The possible user actions include:

<code>answer_unsaved</code>	User makes a change to tokens in the search bar.
<code>answer_saved</code>	User opens an existing saved answer and makes changes to tokens in the search bar.
<code>answer_pinboard_context</code>	User opens an existing saved pinboard, edits a context viz and makes a change to tokens in the search bar.
<code>answer_aggregated_worksheet</code>	User opens an existing saved aggregated worksheet and makes changes to tokens in the search bar.
<code>answer_upgrade</code>	Requests made for the sole purpose of upgrade.
<code>pinboard_view</code>	User opens an existing saved pinboard.
<code>pinboard_filter</code>	User adds, removes or applies values to a pinboard filter.
<code>pinboard_ad_hoc</code>	User drills down in a pinboard viz.
<code>data_chart_config</code>	Request for new data being generated following a chart config change.
<code>data_show_underlying_row</code>	Request to show underlying data for a data row(s).
<code>data_export</code>	Request to export data.

pinboard_tspublic_runtime_filter	Request to TSPublic/pinboarddata with runtime filters.
answer_aggregated_worksheet_save	User updates aggregated worksheet.
answer_add_new_filter	User adds a filter using the UI.
data_show_underlying_viz	Request to show underlying data for a data row(s).
answer_view	User opens an existing, saved answer.
answer_viz_context_view	User opens an existing saved pinboard, edits a context viz.
pinboard_insight_view	User opens SpotIQ tab pinboards.
pinboard_admin_view	User opens admin tab pinboards.
pinboard_embed_view	User opens embed pinboard from a URL.
pinboard_homepage_view	On loading of homepage pinboard.
pinboard_learn_view	On loading learn pinboard.
pinboard_tspublic_no_runtime_filter	Request to TSPublic/pinboard data without runtime filters.

The query underlying this Answer is:

```
top 10
ranked by user actions
user action != 'invalid'
user != {null}
user
last 30 days today
```

The query uses the `TS: BI Server` data source.

Top Pinboards Last Month

This Answer shows the top ThoughtSpot Pinboards ranked by number of views in the last 30 days.

The query underlying this Answer is:

```
top 10
ranked by pinboard views
user action = 'pinboard_embed_view' 'pinboard_tspublic_no_runti
me_filter' 'pinboard_tspublic_runtime_filter' 'pinboard_view'
pinboard != {null}
pinboard
last 30 days today
```

The query uses the TS: BI Server data source.

Ad-hoc search by user

This Answer shows the top ten ThoughtSpot users ranked by the number of ad-hoc searches they have completed.

The query underlying this Answer is:

```
top 10
ad-hoc search
user action = 'answer_pinboard_context' 'answer_saved' 'answe
r_unsaved
by user
sort by ad-hoc search
```

The query uses the TS: BI Server data source.

Today's users

This Answer shows all users who logged in to ThoughtSpot today.

The query underlying this Answer is:

```
latest time  
by user  
user !={null}  
today
```

The query uses the `TS: BI Server` data source.

Configuration Events

This system answer displays recent events that changed the configuration of the system. This panel displays configuration events related to:

Cluster Configuration	Reports configuration actions from the <code>tscli</code> and <code>tql</code> commands.
Metadata Management	Events related to metadata such as column names, column visibility, column and data definition, column rank and so forth.
User Management	Events related to creating, updating, or adding new users and groups.

For a more detailed list, including the user that issued a command, you can use the `tscli event list` command. Administrators can `ssh` into the cluster and specify a time period or even a type of command to include.

```
[admin@testsystem ~]$ tscli event list --since 3d
+-----+
+-----+
|          DATE           |      USER
|          SUMMARY         |
+-----+
+-----+
| 2018-03-06 11:57:10 -0800 PST | eadmin@thoughtspot.int | User
Management: User      |
|                         | "gues
t_1" updated           |
| 2018-03-06 11:48:10 -0800 PST | admin                  | tscl
i node ls               |
| 2018-03-06 11:17:04 -0800 PST | eadmin@thoughtspot.int | Metad
ata Management:        |
|                         | "Metad
ata object "Number of |
|                         | "Oppor
tunity                |
|                         | "AE an
d Stage" of type     |
| ...
| ...
```

Data board

Summary: The Data page shows all the stored tables with details on the last update time, time taken for auto-indexing, number of rows, and so forth.

You can find the Data board under **Data > Usage > Data**.

Table Information		
DATABASE	USER SCHEMA	NAME
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-a6c0991e-462d
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-96a40275-7427
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-bcda2191-cd6c
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-9dc1bfc7-2d27-
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-29546f4d-84ac-
thoughtspot_analytics	falcon_default_sche..	candidates
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-4343525d-261b
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-0694fd57-fadf-
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-51435761-0aac
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-0fb9daec-5230-
dw	falcon_default_sche..	fact_lead_transitions
FalconUserDataDataBase	FalconUserDataSch..	USERDATA-12b3cf23-0de5-
(showing rows 1 - 14 of 913.)		

You can click a column name to sort by table name. This table information is based on an underlying system data. You can present it or copy the link, but you cannot access the underlying query.

Database Status

The **Database Status** column can have the following possible values:

Status	Meaning
READY	The data has been loaded.
IN PROGRESS	The data is still being loaded.
STALE	The data is not up to date.
ERROR	The table is invalid. Call Customer Support.

Replication Status

The **Replicated** column indicates if the table has been replicated or sharded. This can be used in conjunction with the **Total Shards** column to see how your data is distributed.

If the table is replicated on a multi node system, the **Used Capacity (MB)** column will indicate the total space used on all nodes. For example, a 10MB table replicated on a 4 node system will show 40MB used capacity.

Search Status

The **Search Status** column can have the following possible values:

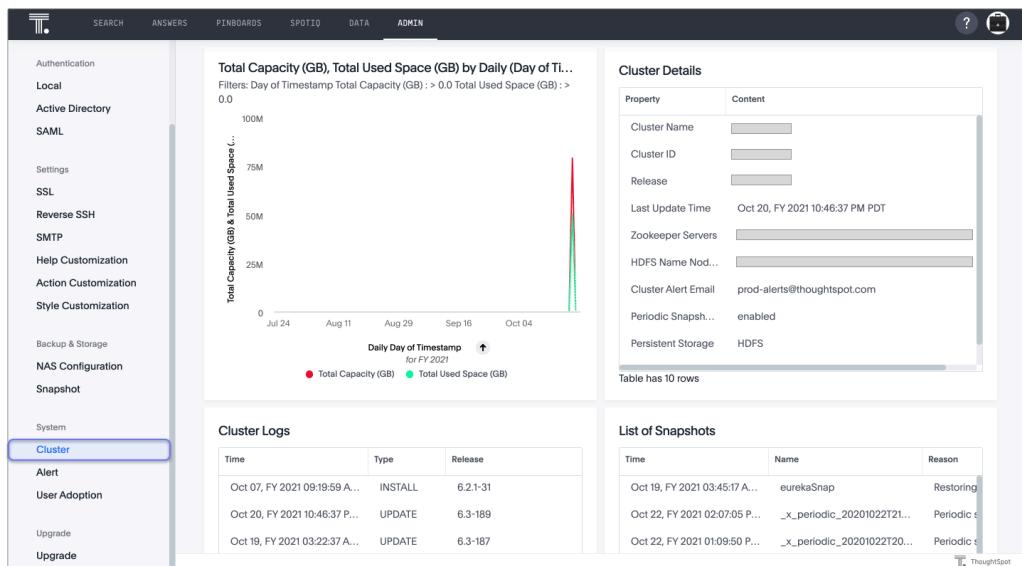
Status	Meaning
READY	The data is up to date and searchable.

Status	Meaning
NOT READY	The data is not ready to be searched.
DELETING INDEX	The table has already been deleted, but the index still exists due to the latency between the database and search engine.
INDEXING DISABLED	Either too many tokens exist in a column for it to be indexed, or indexing has been disabled manually.
CREATING INDEX	The index is being created.
UPDATING INDEX	A change has been made to indexing or the data, and the index is being updated to reflect it.

Cluster Manager board

Summary: Learn about the Cluster Manager board.

You can view cluster information from the Admin Console. The **System Cluster Pinboard** contains several ThoughtSpot Answers that display latency over time, snapshot status, installed release, node functions, and logs. To view this Pinboard, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Cluster** from the side navigation bar that appears.



Only the **Average Latency Last 2 Weeks (sec)** panel contains a standard ThoughtSpot visualization.

The other visualizations rely on internal queries that are not accessible. You can, however, **Present** or **Copy link** on them.

Cluster Details

The fields on this visualization have the following meaning:

Field	Description
Cluster Name	The name of the cluster defined at installation time.
Cluster ID	The ID of the cluster set at installation time.

Last Update Time	Last time the cluster was updated.
Release	Version of the current release.
Zookeeper Servers	IP addresses of the Zookeeper servers.
HDFS Name Nodes	Control nodes for Hadoop Distributed File System (HDFS).

Cluster Logs

The fields on this visualization have the following meaning:

Field	Description
Time	A timestamp indicating when an action occurred.
Type	Type of action.
Release	Identifies the full release number.

You can also use the `tscli logs` command to review log data from your cluster.

List of Snapshots

This visualization shows the snapshots *and the backups* taken on the cluster. The fields on this visualization have the following meaning:

Field	Description
Time	A timestamp indicating when a snapshot or backup happened occurred.
Name	Name of the snapshot file. These files are stored in the `'/usr/local/scaligent/backup'` directory on your cluster.

Reason	Identifies the reason the snapshot/backup was created. You should see several period snapshots if your cluster is configured properly. You may also see evidence here of manual backups. For example, you should be sure your cluster is backed up before major events such as upgrades. Contact ThoughtSpot Support [See page 0] if you don't see evidence your cluster is periodically creating snapshots.
Size	Size of the backup in gigabytes.

Average Latency Last 2 Weeks (sec)

This visualization relies on the `TS: BI Server` worksheet to display the average database latency over the last 15 days. The database latency measures how long it takes for a search to return data from ThoughtSpot - this does not include the time taken to send the answer back to the client, it measures internal processing time. You can use the visualization menu to drill down to its underlying query:

```
average datacache (sec)
average total (sec)
daily
last 15 days
for database latency (us) > 0
```

Related information

`tscli logs` command [See page 0]

Alerts and Events board

Summary: The Alerts and Events section shows notifications, alerts, and an audit trail of cluster configuration changes.

You can view alert information for your ThoughtSpot cluster from the Admin Console. The **System Alert Pinboard** contains information on alerts, configuration events, and notification events. To view this Pinboard, navigate to the Admin Console by clicking on the **Admin** tab from the top navigation bar. Select **Alert** from the side navigation bar that appears.

Time	Type	Message
Jul 23, 2020 09:10:06 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:09:52 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:09:38 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:09:25 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:09:11 PM U...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:08:58 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:08:44 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:08:30 PM ...	TASK_TERMINAT...	Task falcon.worker tr...
Jul 23, 2020 09:07:57 PM ...	TASK_TERMINAT...	Task falcon.worker tr...

Table has 713 rows

Time	User	Summary
Jul 23, 2020 08:59:24 PM ...	tsadmin	User Management: I...
Jul 23, 2020 08:58:11 PM U...	admin	tscli service add-jav...
Jul 23, 2020 08:57:48 PM ...	admin	tscli service add-jav...
Jul 23, 2020 08:40:23 PM ...	tsadmin	User Management: I...
Jul 23, 2020 08:40:08 PM ...	tsadmin	User Management: I...
Jul 23, 2020 08:38:23 PM ...	admin	tscli service add-jav...
Jul 23, 2020 12:48:20 AM ...	tsadmin	User Management: I...
Jul 23, 2020 12:47:41 AM U...	tsadmin	User Management: I...
Jul 23, 2020 12:45:46 AM ...	paul@thoughtsp...	User Management: I...

Table has 72 rows

Time	User	Summary

This Pinboard contains three ThoughtSpot Answers: **Alerts**, **Configuration Events**, and **Notification Events**. You can **present** these Answers or copy their embed link.

Alerts

The fields on this Answer have the following meaning:

Field	Description
Time	When the alert was sent.
Type	The ID of the event.

Message	The text of the alert message.
----------------	--------------------------------

For a full reference of possible alerts, see the [Alert code reference \[See page 0\]](#).

Configuration Events

This system answer displays recent events that changed the configuration of the system. This answer displays the **Time**, the **User** that performed the action, and a **Summary** of the action.

Notification events

This answer displays notifications of data loads. The answer displays the **Time**, the **User** that performed the action, and a **Summary** of the action. Notifications are kept for 90 days before being discarded.

System worksheets

Summary: Learn about the system worksheets that ThoughtSpot provides.

Most of the monitoring information in **System Health** comes from system worksheet which administrators can view, but not modify. The underlying tables are protected system tables that cannot be accessed directly. However, administrators can create new, custom monitoring reports from the worksheets.

List the system worksheets

To list the system worksheets:

1. Go to the **Data** tab.
2. Choose **All** and **Worksheets**.
3. Enter `TS:` in the search field.

The screenshot shows the ThoughtSpot interface with the 'Data' tab selected. In the top navigation bar, the 'Data' tab is highlighted. Below the navigation, there are tabs for 'Tables' and 'Data Sources'. Underneath these, there are filters for 'All', 'Yours', 'All types', and 'Worksheets', with 'Worksheets' being the active filter. A search bar contains the text 'Q TS:'. The main area displays a table with four columns: 'Name', 'Source', 'Stickers', and 'Materialize Status'. The table lists four system worksheets: 'TS: BI Server', 'TS: Search', 'TS: Database', and 'TS: Service Resources'. Each row includes a timestamp ('Modified') and a user icon ('System User').

Name	Source	Stickers	Materialize Status
TS: BI Server			11 months ago
TS: Search			a year ago
TS: Database			2 years ago
TS: Service Resources			2 years ago

Summary of the worksheets

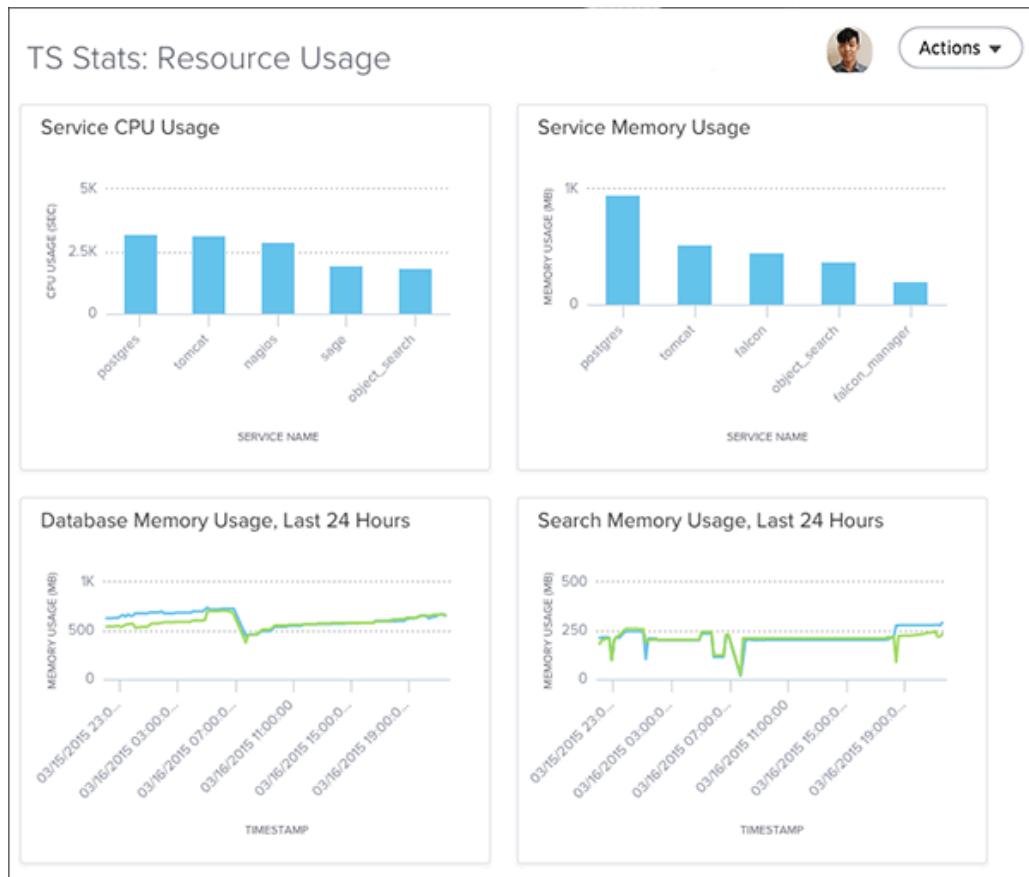
Worksheet	Description
-----------	-------------

TS: BI Server	Contains data related to the systems associated with underlying the ThoughtSpot BI server. This includes database latency, browser clients, size of responses, and more.
TS: Database	Contains information related to the database cache and queries run on the database. For example, you could use this worksheet to see data on the query errors returned by the database.
TS: Internal Capacity WS	Describes cluster memory capacity by node over time.
TS: Internal Table Wise Capacity WS	Describes memory capacity by node, table name, shard count, and CSV replication over time.
TS: Metrics	Contains metrics by cluster and host over time.
TS: Search	Contains data related to the number of searches (queries) run in the system. This contains information such as uptime, host, and timestamps.
TS: Service Resources	Contains data related to cluster nodes including page faults, memory usage, memory failures, and more.
TS: Table Info	Describes the named tables by timestamp, row count, row shards, and row size.
TS: Table Row Counts	Describes the named tables by timestamp and row count.
TS: Table Row Counts and Shards	Describes the named tables by timestamp, row count, row shards, and row size.
TS: Table Row Size	Describes the named tables by timestamp and row size.
TS: Table Shards	Describes the named tables by timestamp and shard count.

System pinboards

Summary: ThoughtSpot provides several pinboards that help with system monitoring.

There are several system monitoring pinboards in ThoughtSpot that provide answers for system status and resource usage questions. The information in these pinboards are updated hourly from internal data sources that collect monitoring statistics.



Only users with administrative privilege can view the monitoring pinboards. They are based on worksheets, which administrators can view, but not modify. However, you can create new monitoring pinboards from the worksheets.

View system pinboards

To view these system pinboards:

1. Go to the **Pinboards** tab.
2. Choose **All**.
3. Search for the Pinboard you would like to view. You can view the **System Cluster Pinboard**, the **System Alert Pinboard**, and the **User Adoption Pinboard** from the Admin Console.

Summary of system pinboards

Pinboard	Description
Learn how to use ThoughtSpot	Contains search replays created in the system.
TS Stats: Alert Detail	Combines alerts, notification events, and configuration events boards.
TS Stats: Cluster Detail	Contains cluster details, logs, snapshots, and latency data.
TS Stats: System Information and Usage	Replicates the Admin > System Health > Overview page.
TS Stats: Latency Visualizations	Latency on servers and impression [See page 613] counts. This data comes from the TS: BI Server worksheet.
TS Stats: Table Status	Replicates the Data board.
User Adoption	Contains information on how your ThoughtSpot users are interacting with ThoughtSpot, and how your user adoption is changing over time. Refer to User Adoption Pinboard [See page 75] .
Performance Tracking	Contains information on how your ThoughtSpot cluster is performing. Refer to Performance Tracking Pinboard [See page 621] .

About deprecated boards

The deprecated boards (`TS status: Usage - Deprecated` and `TS StaTS: Queries -- Deprecated`) are there to support older installations. New installations should not use or rely on deprecated boards. Older installations that have used these boards in some way should use the new boards and remove any dependencies.

Falcon Monitor Pinboards

Summary: Use the Falcon Monitor Pinboards for an overview of Falcon, ThoughtSpot's in-memory database, and its health, based on query, data load, system stats, and varz metrics.

In ThoughtSpot release 6.2, there are 5 new Pinboards, based on Falcon metrics, that are available to system administrators.

Falcon is ThoughtSpot's in-memory database. Falcon monitoring functionality pushes different kinds of metrics to Falcon system tables every fifteen minutes. These system tables, when updated, update the 5 new Pinboards that you can use to monitor Falcon's health. You can see these Pinboards from the **Pinboards** page, by searching for **360_Overview**:

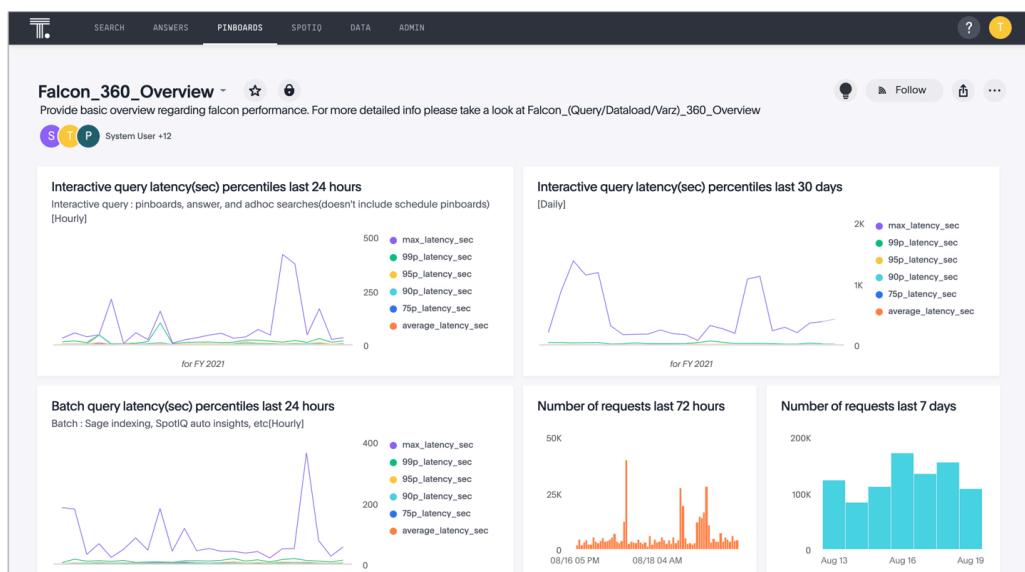
Name	Stickers	Modified ↓	Author
Falcon_360_Overview Provide basic overview regarding falcon performance. For more detailed info ple...		a month ago	System User
Falcon_Query_360_Overview Provides info about Falcon query execution based on traces		a month ago	System User
System_Stats_360_Overview Provides system stats info such as memory usage, cpu usage, load avg, etc		a month ago	System User
Falcon_Dataload_360_Overview Provides info about Falcon dataloads based on traces		2 months ago	System User
Falcon_Varz_360_Overview Provides info about Falcon services based on varzs		2 months ago	System User

The 5 new database monitoring Pinboards are: [Falcon_360_Overview \[See page 617\]](#), [Falcon_Query_360_Overview \[See page 617\]](#), [Falcon_Dataload_360_Overview \[See page 618\]](#), [Falcon_Varz_360_Overview \[See page 619\]](#), and [System_Stats_360_Overview \[See page 619\]](#).

You can use these Pinboards for proactive monitoring, or, with help from [ThoughtSpot Support \[See page 0\]](#), for debugging.

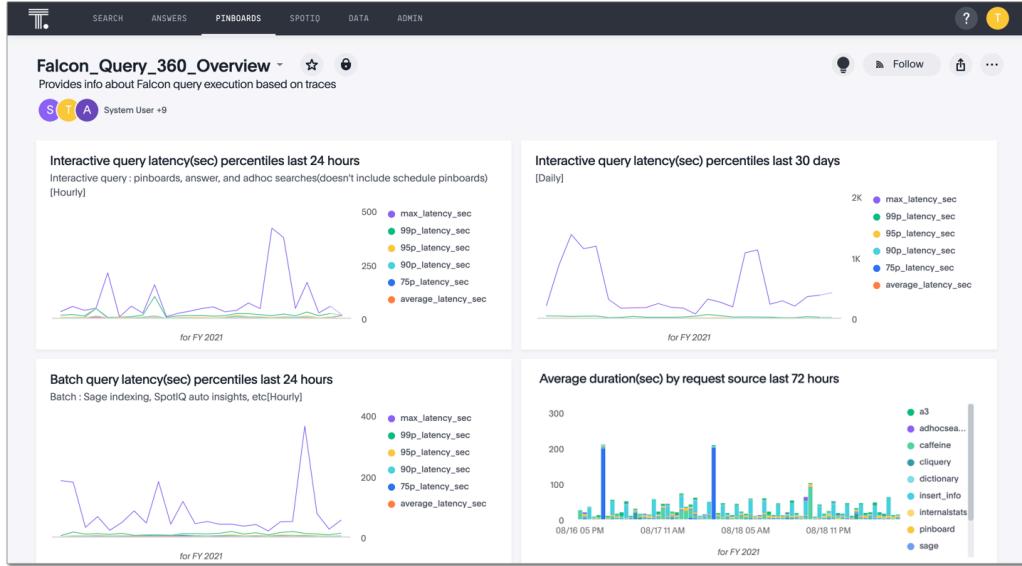
Falcon_360_Overview

This Pinboard provides basic information regarding Falcon's performance and health. Visualizations include **Interactive query latency(sec) percentiles last 24 hours**, **Dataloader : Avg Ingestion Speed (# Rows Ingested / Load Time) Per Hour - Last 72 Hours**, **CPU Utilization (System, Idle, User) - Last 72 Hours**, **Dataloader Frequency By Hour Of Day (Aggregated over 7 days)**, **Top 10 frequently changed pinboard vizes**, and so on.



Falcon_Query_360_Overview

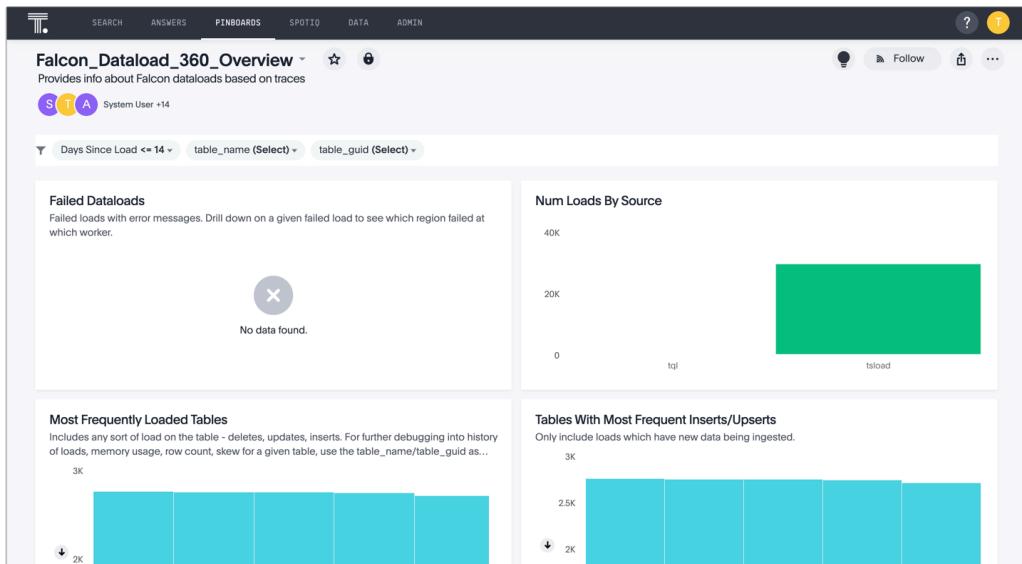
This Pinboard provides information about Falcon query execution based on traces. Visualizations include **Interactive query latency(sec) percentiles last 24 hours**, **Average duration(sec) by request source last 72 hours**, **Median latency(sec) by hour of the day last 7 days**, **Count of trace ids by error status last 72 hours**, **Max JIT compilation time(sec) last 72 hours**, **Top 10 vizes based on avg duration(sec)**, and so on.



Falcon_Dataload_360_Overview

This Pinboard provides information about Falcon data loads based on traces. Visualizations include

Failed Dataloads, Tables With Most Frequent Inserts/Upserts, Table Growth (# Rows) Over Time, Load Frequency By Hour Of Day (Aggregated in a time window), Slowest Loads & Corresponding Region Load Time Skew, Loads With Highest Compaction Overhead (# Rows) and so on.

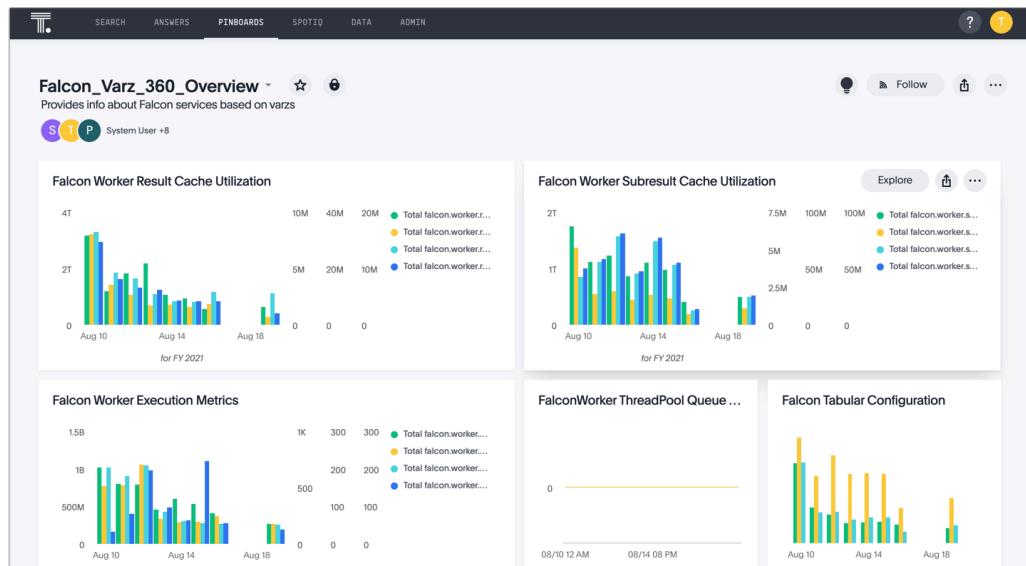


Falcon_Varz_360_Overview

This Pinboard provides information about Falcon services based on metrics in VarZ format.

Visualizations include **Falcon Worker Execution Metrics**, **Falcon Query Runtime (Average and Max)**,

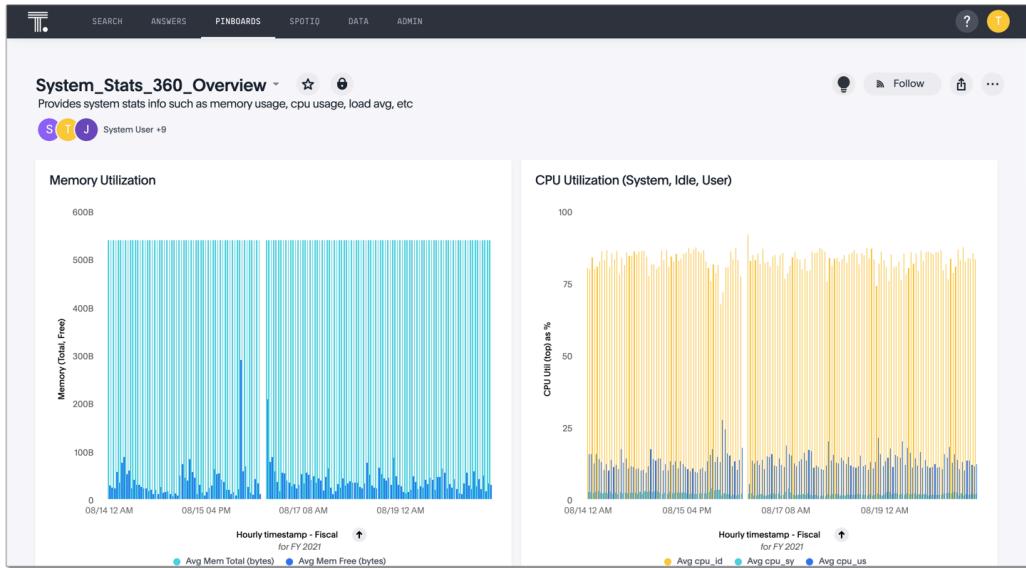
Falcon Worker Memory Manager, **Daily Data Load Statistics**, **Falcon Compiler Cache Daily Usage**, and so on.



System_Stats_360_Overview

This Pinboard provides system stats information such as memory usage, CPU usage, and load average.

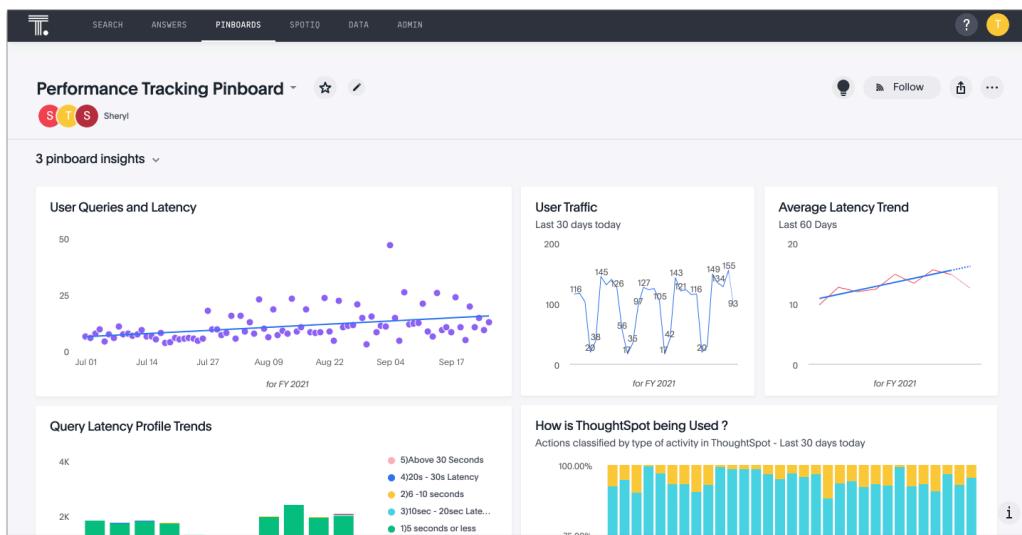
Visualizations include **Memory Utilization**, **CPU Utilization (System, Idle, User)**, **Loadavg (current/total scheduling entities)**, **Memory (Free)**, **Netstat connections**, and so on.



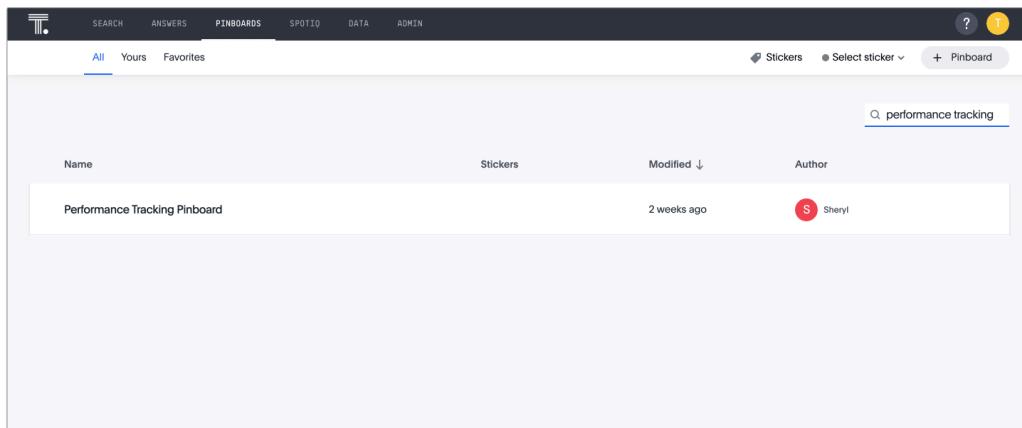
Performance Tracking Pinboard

Summary: Use the Performance Tracking Pinboard to understand how your ThoughtSpot cluster is performing.

The Performance Tracking Pinboard contains essential information about your ThoughtSpot cluster's performance. Use this Pinboard to understand query latency, user traffic, how users are interacting with ThoughtSpot, most frequently loaded tables, and so on.



To view this Pinboard, click on the **Pinboards** tab from the top navigation bar. Search for **Performance Tracking**.



Use this Pinboard to view and analyze information about the following topics:

- query latency
- whether users are searchers or consumers
- user traffic
- database latency
- query timeouts
- frequently loaded tables
- data requests and latency
- data errors
- pinboard latency

About troubleshooting

Summary: Learn the basics of troubleshooting for ThoughtSpot.

The information here provides very basic troubleshooting.

- **[Get your configuration and logs \[See page 625\]](#)**

For troubleshooting on specific incidents or cluster problems, getting a log bundle can help.

- **[Upload logs to ThoughtSpot Support \[See page 630\]](#)**

- **[Network connectivity issues \[See page 632\]](#)**

If network connectivity to and from ThoughtSpot is not working, try using these steps to find and correct the issue.

- **[Check the timezone \[See page 633\]](#)**

ThoughtSpot comes configured with the timezone where it is to be installed.

- **[Browser untrusted connection error \[See page 634\]](#)**

If you are not using a SSL certificate for authentication, users will see an untrusted connection error in their browser when accessing ThoughtSpot. The error looks slightly different depending upon the Web browser being used.

- **[Characters not displaying correctly \[See page 635\]](#)**

Your CSV files are more likely to load smoothly if they are encoded with UTF-8. If you're having problems with some characters rendering incorrectly, you can convert the files to UTF-8 encoding before loading the data.

- **[Clear the browser cache \[See page 636\]](#)**

You might occasionally see unexpected behavior that is due to the Web browser caching information from ThoughtSpot. In this case, clearing the browser cache and reloading the page should resolve the problem.

- **[Cannot open a saved answer that contains a formula \[See page 639\]](#)**

- **[Data loading too slowly \[See page 642\]](#)**

Some tables may take an unusually long time to load due to a high data version issue. This issue normally arises when ThoughtSpot completes an upgrade or the system is recovering after a crash.

- **[Search results contain too many blanks \[See page 643\]](#)**

If you find that your search results contain too many blanks when your data source is a worksheet, there is a simple adjustment you can make to fix this.

About troubleshooting

For more detailed troubleshooting, [Contact ThoughtSpot \[See page 0\]](#).

Get your configuration and logs

Summary: Two main troubleshooting tools are getting a log bundle and understanding your cluster configuration.

For troubleshooting on specific incidents or cluster problems, two things are important: understanding your current configuration and getting a log bundle.

Check your configuration

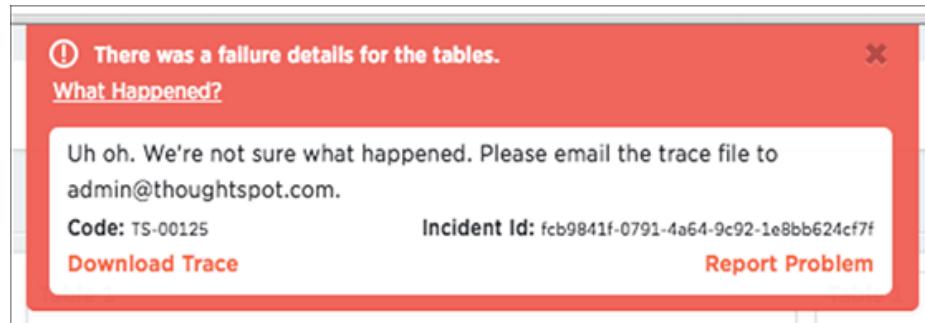
1. Log in to the ThoughtSpot cluster as the `admin` user.
2. Use the `tscli feature` subcommand to display your current configuration.

ACTION	NAME	STATUS	CONFIGUR
	Firewall	Disabled	
	Saml	Disabled	
	Ldap	Disabled	
	CustomBranding	Disabled	
	CustomBrandingFontCustomization	Disabled	
	DataConnect	Disabled	
	RLS	Enabled	
	Callhome	Enabled	
	SSHTunnel	Enabled	
	Fileserver	Disabled	

How to get logs

There are two ways to get logs:

- When ThoughtSpot encounters a problem, a red bar displays in the browser with an error message. You can click **What Happened?** in the error message for more details. To download related logs, click **Download Trace**. Send the logs as an email attachment to the support contact that is provided. Clicking **Report Problem** will also send the logs as an email attachment to your administrator.



- You can generate a log bundle using the `tscli` command `tscli logs collect` if you are comfortable with Linux. The command lets you specify which logs to collect and from what time periods.

Usage for this command is:

```
tscli logs collect
  --include <selector | glob>
  [--exclude <selector | glob>]
  [--since <hours,minutes,days>
   | --from <yyyymmdd-HH:MM>
   --to <yyyymmdd-HH:MM>]
  [--out <path>]
  [--maxsize <size_in_MB_or_GB>]
  [--sizeonly]
```

The full list of all selectors is:

- `all` collects all of the logs listed from the system and the ThoughtSpot application.
 - `system` collects all system logs, e.g. syslog, upstart, mail logs, etc.
 - `ts` collects all logs from the ThoughtSpot application. This includes falcon, sage, orion core (cluster management), etc.
 - `orion` collects all orion logs including cluster management, hdfs, zookeeper, etc.
- Detailed syntax and options are listed in the [tscli command reference \[See page 0\]](#).

Examples

Here are some examples of usage for the command `tscli logs collect`:

To collect all logs from the past day to the default location (`/tmp/logs.tar.gz`):

```
$ tscli logs collect --include all --since 1d
```

In this example, `all` is a selector for all the available logs.

In most cases, you can probably use the selector `ts` to only capture logs for the ThoughtSpot application:

```
$ tscli logs collect --include ts --since 2d
```

For debugging cluster management issues, use a command like this one, which collects logs for system and orion from the past 2 hours. The output is written to `/tmp/debug.tar.gz` as specified using `--out`:

```
$ tscli logs collect --include system,orion --since 2h --out /tmp/debug.tar.gz
```

This command collects logs from a specific time window:

```
$ tscli logs collect --include system,orion --from 20150520-12:00 --to 20150522-12:30
```

Advanced usage alert! You can also use `--include` and `--exclude` to specify filesystem paths as a glob pattern. This works like the Linux `find(1)` command. Pass all the entries in `--include` starting with `/` to `find(1)`, and all entries in `--exclude` which are not selectors to `find(1)` using the `-not -path` flag.

```
$ tscli logs collect --include system,orion --exclude *hadoop*,*zookeeper* --since 2h
```

The above command collects all system and all orion logs, but excludes hadoop (hdfs) and zookeeper logs. See [Upload logs to ThoughtSpot Support \[See page 630\]](#) about using a secure file sever to collect log files or other files needed for troubleshooting. You can easily send log files to this file serve directly from the ThoughtSpot instance.

Upload logs to ThoughtSpot Support

Summary: You can send logs to ThoughtSpot Support to get troubleshooting help.

ThoughtSpot Support uses a secure file sever to collect log files or other files needed for troubleshooting. You can easily send log files to this file server directly from the ThoughtSpot instance.

Metrics collection

ThoughtSpot collects the diagnostic information from your system on an ongoing basis: there is no time needed to collect diagnostic information after a problem is reported. These metrics allow for:

- Our support team can begin working to remediate any issue with you immediately.
- Metrics provides direct visibility to the ThoughtSpot team on your system's limits. Therefore, our Support team can proactively identify critical threshold issues and work to prevent failures. Metrics also help reduce SLA times as the team can debug much faster.
- ThoughtSpot can tune search algorithms by studying search history and schema.
- ThoughtSpot analyzes expensive and complex query patterns to look for performance optimizations.

Finally, the metrics pipeline allows ThoughtSpot to identify application-use patterns that contribute to performance bottlenecks with specific browsers and help your team prevent or alleviate them.

Other log uploads

Before you can upload a file to the secure file server:

1. [Configure the connection to the file server \[See page 139\]](#).
2. Obtain the directory path on the file server.

The server directory path for uploading a file is formatted like this example: /Shared/
support/<customer_name>. If you do not know the customer name, [contact ThoughtSpot Support \[See page 0\]](#).

You can upload files directly to the file server using this procedure:

1. Log in to the Linux shell using SSH.
2. Navigate to the directory where the file to be uploaded is located.
3. Issue the command to upload the file, specifying the file name and directory path:

```
$ tscli fileservice upload --file_name <file> --server_directory_path <path>
```

When your upload succeeds, you will see a confirmation message.

Network connectivity issues

Summary: If network connectivity to and from ThoughtSpot is not working, try using these steps to find and correct the issue.

To troubleshoot network connectivity for ThoughtSpot:

1. Make sure that the network cables are connected correctly.
2. Check that the network cable is connecting the nodes to the network switch.
3. Try replacing the cable with a cable from a known working system to rule out a bad cable or switch connectivity issues.
4. Make sure the eth0 interface is connected to the network by issuing: `ethtool eth0` The port that's currently connected will have "link detected" in the last line of the output.
5. If the networking settings have been reconfigured, reboot each of the nodes.

Check the timezone

Summary: Learn how to check the timezone your ThoughtSpot installation is running on.

ThoughtSpot comes configured with the timezone where it is to be installed. Data is imported based on the timezone of the node from which `tsload` or `tql` is run. To see the timezone your ThoughtSpot node is running under, log in to the server as the `thoughtspot` user and run the `date` command:

```
[thoughtspot@ts.server etc]$ date  
Tue Feb 20 09:07:04 PST 2018
```

To see the current timezone used by the ThoughtSpot application, choose **Admin > Cluster Manager** and review the **Cluster Details**:

Cluster Details	
Property	Content
Cluster Name	local
Cluster ID	local
Release	dev
Last Update Time	Feb 16, 2018, 4:14:31 AM PST
Zookeeper Servers	172.18.248.8:2181
HDFS Name Nodes	172.18.248.8:8020

The timezones should match.

Sometimes the timezone that is listed by `date` is not the active timezone for the ThoughtSpot application and the application needs to be reset. If you need to change the timezone, [contact ThoughtSpot Support \[See page 0\]](#) and they will change the timezone for you. You may need to change the timezone if you move the server between data centers.

Browser untrusted connection error

Summary: If you do not have an SSL certificate, users get an untrusted connection error.

If you are not using an SSL certificate for authentication, users will see an untrusted connection error in their browser when accessing ThoughtSpot. The error looks slightly different depending upon the Web browser being used.

ThoughtSpot uses secure HTTP (the HTTPS protocol) for communication between the browser and ThoughtSpot. By default there is no SSL certificate for authentication. This must be added by the site administrator. If the site administrator has not added the certificate, the browser warns the user.

Browser	Warning
Google Chrome	The site's security certificate is not trusted!
Mozilla Firefox	This Connection is Untrusted

If you see the warning message, choose one of the following options:

- Ask the site administrator to install the certificate.
- Ask the site administrator to turn off SSL using this command in the shell on the ThoughtSpot instance:

```
$ tscli ssl off
```

- You can choose to ignore the message, and access ThoughtSpot without SSL.

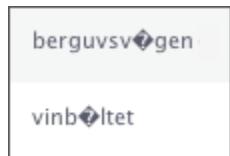
Characters not displaying correctly

Summary: If characters in your data do not display correctly in ThoughtSpot, make sure they are UTF-8 encoded.

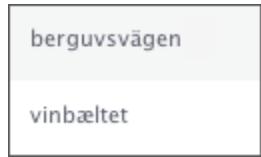
Your CSV files are more likely to load smoothly if they are encoded with UTF-8. If you're having problems with some characters rendering incorrectly, you can convert the files to UTF-8 encoding before loading the data.

You might see unexpected characters in your data, especially characters whose ASCII values are at the high and low end of possible values. Some examples of characters that can appear incorrectly are: æ, ñ, ä, í, ö.

If this happens, your data will look like this:



Instead of displaying correctly like this:



To encode your data as UTF-8:

1. On Windows, open your CSV file in Notepad. Save the file as CSV with the Unicode option.
2. On Linux or MacOS, issue a command like:

```
$ iconv -f -t UTF-8 <in_file>.csv > <out_file>.csv
```

3. Reload the data.
4. Attempt to import it again.

Clear the browser cache

Summary: Clear the browser cache if you have unexpected network issues.

You might occasionally see unexpected behavior that is due to the Web browser caching information from ThoughtSpot. In this case, clearing the browser cache and reloading the page should resolve the problem. You can usually resolve these situations by clearing the browser cache:

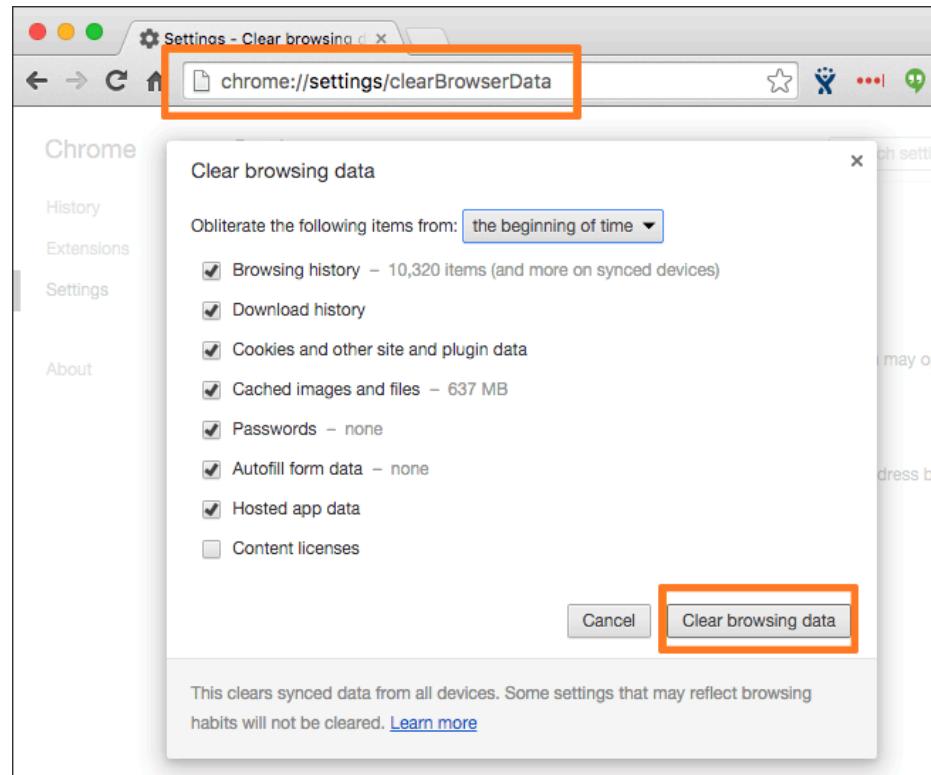
- During a ThoughtSpot session, the browser suddenly displays a white screen and reloading does not fix the problem. This is due to a self-signed SSL certificate that has timed out during the session.
- When accessing the Help Center, you see a login screen. This is due to a problem during automatic authentication in the Help Center, after which the bad login gets cached by the browser.

To resolve any of these situations, clear the browser cache:

1. Clear the browser cache.

This works a little differently on individual browser versions and operating systems. For example, when using Chrome, to get to the browser cache settings, navigate to:

```
chrome://settings/clearBrowserData
```

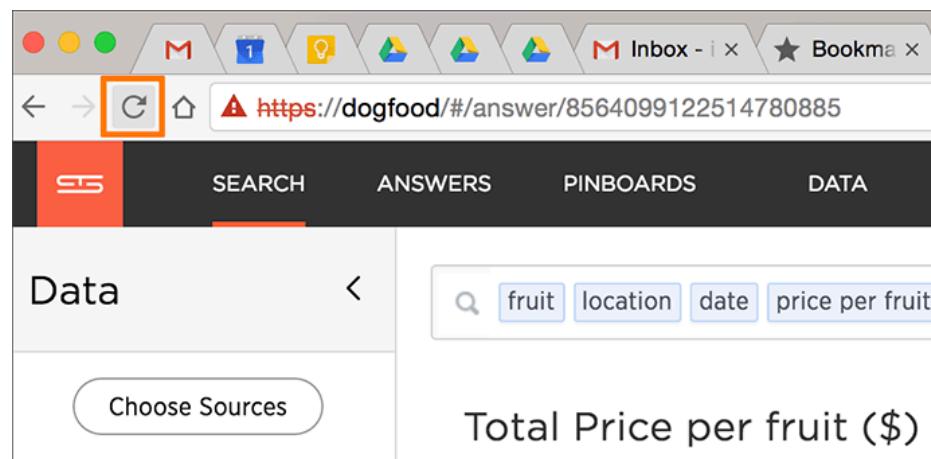


2. Click **Clear browsing data**.

This is the name of the button on Chrome. The name may vary slightly on other browsers.

3. Reload the page.

For example, on Chrome you would click the **Reload** icon:



Clear the browser cache

Now the problem should be fixed, and the page will appear as expected.

Cannot open a saved answer that contains a formula

Summary: Learn how to troubleshoot problems with data types and formulas.

When working with formulas, keep in mind the data types they return. You may occasionally see unexpected results, or even be unable to open a saved answer, due to problems with data types and formulas.

In this scenario, “data type” refers the data type as defined in the column definition when creating the schema (INT, TIMESTAMP, VARCHAR, etc.).

When you define a formula, the data type it returns is set automatically. This can lead to problems, if you build another formula that uses the output of the first formula as input. This can be hard to understand, so an example will be helpful.

Suppose you have created a worksheet that contains a formula called “weekday” defined as:

```
day_of_week(date)
```

The output of that formula is the day of the week (Monday, Tuesday, etc.) returned as a text string (VARCHAR, ATTRIBUTE).

Then suppose you create an answer using the worksheet as a source. And in the answer, you create another formula on top of the formula column in the worksheet. This formula is supposed to return the day of the week that is two days after the given day of the week:

```
weekday + 2
```

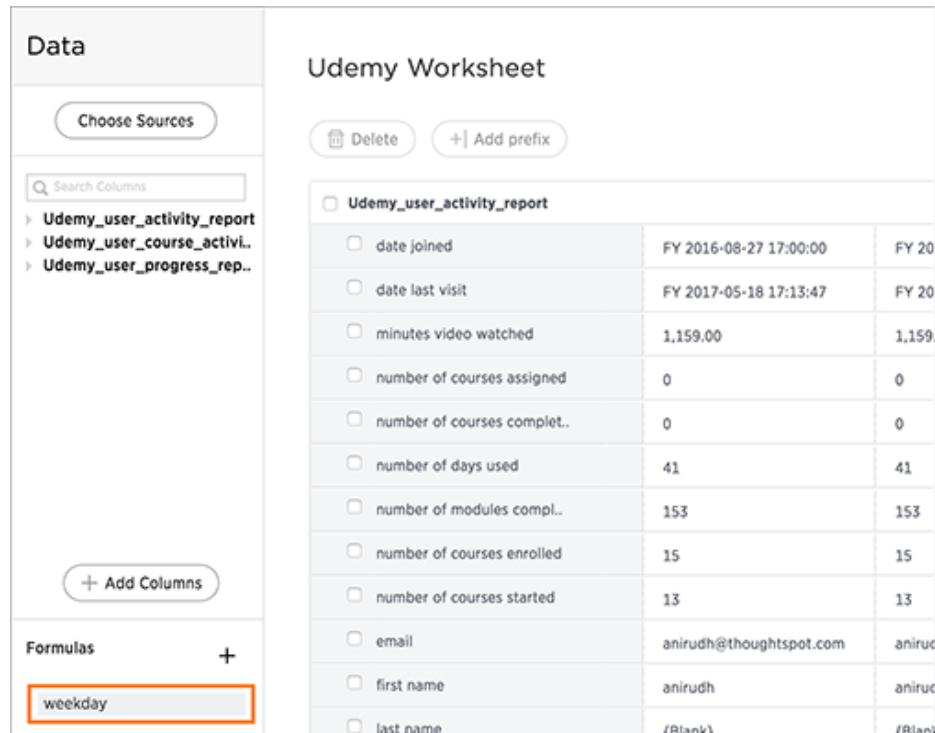
In this case, you have effectively created a formula on top of another formula. This works fine, so long as the data types in the worksheet formula can work in the answer formula. If not, you may not be able to save the answer, or open it after it has been saved. Here, the second formula you created does not work, because it is invalid. It is trying to subtract a number from a text string.

If you encounter this issue, you must open the worksheet and edit its formula so that it returns the type expected by the formula that was built on top of it. In this case, a numeric data type.

You must change the underlying worksheet column to use day_number_of_week instead of day_of_week. This is because day_number_of_week returns a numeric data type.

Here are the steps to resolve an issue like this:

1. Open the underlying worksheet that contains the formula whose output data type you need to change.
2. Click the formula name to edit the formula.

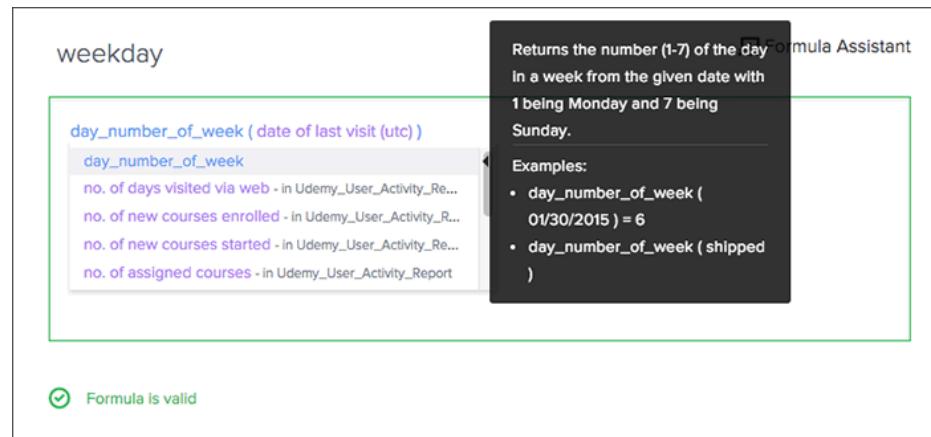


The screenshot shows the ThoughtSpot Data interface. On the left, there's a sidebar with 'Data' at the top, followed by 'Choose Sources' and a search bar labeled 'Search Columns'. Below that is a list of columns: 'Udemy_user_activity_report', 'Udemy_user_course_activi..', and 'Udemy_user_progress_rep..'. At the bottom of this sidebar are 'Add Columns' and 'Formulas' sections. The 'Formulas' section contains a single entry: 'weekday'. This entry is highlighted with a red box. To the right of the sidebar is the main workspace titled 'Udemy Worksheet'. It features a toolbar with 'Delete' and 'Add prefix' buttons. Below the toolbar is a table with the following data:

Udemy_user_activity_report		
<input type="checkbox"/> date joined	FY 2016-08-27 17:00:00	FY 20
<input type="checkbox"/> date last visit	FY 2017-05-18 17:13:47	FY 20
<input type="checkbox"/> minutes video watched	1,159.00	1,159.
<input type="checkbox"/> number of courses assigned	0	0
<input type="checkbox"/> number of courses complet..	0	0
<input type="checkbox"/> number of days used	41	41
<input type="checkbox"/> number of modules compl..	153	153
<input type="checkbox"/> number of courses enrolled	15	15
<input type="checkbox"/> number of courses started	13	13
<input type="checkbox"/> email	anirudh@thoughtspot.com	anirudh
<input type="checkbox"/> first name	anirudh	anirudh
<input type="checkbox"/> last name	{Blank}	{Blank}

3. In the Formula Builder, modify the formula, so that it returns the expected data type.

There are data type conversion formulas available to make this easier. To view them and their syntax, open the **Formula Assistant**, and expand the section called **Conversion**.



4. Make your changes, and saving the formula by clicking **Save**.
5. Save the worksheet by clicking **Save**.
6. Now you will be able to open the answer that was created on top of the worksheet.

Data loading too slowly

Summary: If your data is loading slowly, there are a few things you can do to fix it.

Some tables may take an unusually long time to load due to a high data version issue. This issue normally arises when ThoughtSpot completes an upgrade or the system is recovering after a crash.

The data version is the number of loads that have been historically applied to a table. Every completed load increments the version number of the table by one. ThoughtSpot would need to process each version of the table during restoration, which could increase the time it takes to build the table.

There are a few steps you can take to check for a high data version issue and fix it. To improve data loading speed, run the following command to find the number of tables that are building and their names

```
tscli cluster status --mode table
```

You may notice that a few small tables are taking up a lot of time to be built. This could simply be due to the deceptively large size of the table. There is also the chance it could be due to a high data version issue. To determine if this is a high data version issue, check the size of the table by running the following command:

```
echo 'show statistics for server;' | tql
```

If there is a large number of rows in the table, go on to shard the table. If the table has a small number of rows, then the slow loading speed is caused by a high data version issue, and you do not have to shard the table. Use the compact table functionality to trim the table down to its actual size:

```
tql> compact table <table name>;
```

Search results contain too many blanks

Summary: Learn how to fix the problem when your search results contain too many blanks.

If you find that your search results contain too many blanks when your data source is a worksheet, there is a simple adjustment you can make to fix this.

If you find that the charts and tables built on a worksheet contain a large number of null values (which display as {blank} in the web browser), you can fix this by changing the [inclusion rule \[See page 0\]](#) for the worksheet.

An inclusion rule that specifies **Exclude Empty Rows (Inner Join)** will reduce the number of null values in the result. Imagine a worksheet that includes data about a retail grocery store. There are rows in the worksheet from these source tables:

Table Name	Description
sales	Fact table with sales made per product per store.
products	Dimension table with information about every product.
stores	Dimension table with information about every store.

When creating the worksheet, suppose you choose **Include Empty Rows (Left Outer Join)** for the inclusion rule and **Progressive Joins** for the join rule. In this case, if you type “product name” in your search, you can see a list of all the products that exist. Suppose you then add “store name” to your search. You will see a lot of null ({blank}) values in the result. This happens because the columns “store name” and “product name” are joined through the fact table, “sales”. So for every product that has never been sold in a particular store, you can see {blank} in the “store name” column. This may be what you want to see, in which case, you can leave the worksheet as is, and choose **Exclude** for the {blank} values in your table or chart, whenever you don’t want to see them.

However, in many cases, including all the {blank} values could confuse end users. So if you encounter this problem, you can [edit the worksheet, and change the inclusion rule \[See page 481\]](#) to **Exclude Empty**

Rows (Inner Join). Now when searching for “store name” and “product name” on the worksheet, users will not be overwhelmed by null values. They’ll only see the rows where a particular product has been sold in a particular store.