# Spotfire Statistica®

## Security Guide

Version 14.2.0 | March 2024

# Contents

# Introduction

The goal of a data science product like Spotfire Statistica® is to access data and generate analytic results.

The raw data might contain personal identifying information and can be a potential liability due to privacy laws, like General Data Protection Regulation (GDPR). Or the raw data can be seen as an important investment that needs to be protected from corporate spying. The analytic results can be sensitive because it is the special feature with which the company can keep operational costs under control, maximize a sales plan, find fraud, discover a drilling site, and perform drug discovery.

To accommodate the security concerns associated with these factors, Statistica implements role-based access controls (RBAC) for its centralized metadata store, also known as Statistica Enterprise database. It supports authentication based on Windows Active Directory or through a custom user account database. Authorization is achieved through roles and groups defined in the metadata store, controlling access and starting specific applications in the software suite, as well as the ability for defining and using metadata objects like database connections, queries, and analyses.

These metadata (objects) are created, edited, reviewed, versioned, approved, and deleted using the Statistica Enterprise Manager application, which is a part of the Statistica client installation.

# Product Connectivity - Authentication

For customers who own Spotfire Statistica® Server or Spotfire® Data Science - Operations, the following applies:

The Statistica client application authenticates when started.

The application uses either Windows Integrated Authentication, using the current Windows user identity or an explicit login with account credentials defined in the Statistica Enterprise user database. Statistica client application has two utility components to retrieve data from a database, Statistica Query and Advanced Query. The user needs an additional login and password to be authorized to retrieve data.

One of the metadata elements is users and groups for roles based security. These users and groups can be synced with Windows domain groups or can be synced with local users on the server. Or custom user accounts can be created within the Statistica system.

Here are example group names that define what applications work with the login. A user can be listed in one or all of the groups, depending on their role. These are system groups that grant access to start an application or log in to a web server. They do not actually grant permission to execute a specific analytic project.

- Ad-hoc Analysis; user can log in to the Statistica application with USR permission
- Administrators; user can log in to the Statistica Enterprise Manager application with SADM permission
- Manual Data Entry Users; user can log in to Statistica Data Entry web server with DE permission
- Spotfire Users; Spotfire Analyst or Spotfire Consumer open DXP file that needs permission to execute a Workspace with WUSR permission
- Web Users; user can log in to WebStatistica with WUSR permission

For more information, see the Server Administrator Guide.

# Spotfire® Service for Statistica Authentication

Spotfire® Service for Statistica is also known as Statistica Service. Read the *Spotfire® Service for Statistica Installation and Configuration Guide* prior to reading this section.

## Statistica Service Configuration

This section is focused on the following properties: `statistica.enterprise`, `statistica.enterprise.user`, and `statistica.enterprise.password`.

The `statistica.enterprise` property must be set to TRUE to allow Spotfire consumers to execute Spotfire Analysis (DXP files) that are linked to Statistica Workspaces (SDM files) saved in Statistica Enterprise database. An organization can use this configuration and access the following capabilities:

- If you need an audit trail of every execution for a Statistica Workspace linked to a specific user, date, and time. This information is captured within Statistica audit log with this configuration.
- Linking, rather than embedding, the SDM file to the DXP file potentially means fewer steps to publish changes. The SDM file can be modified and governed within Statistica.
- Data scientists can create one Workspace and use it within many DXP files. This makes governance easier with fewer objects to manage and approve.

If Enterprise integration is enabled with the statistica.enterprise switch, administrative Statistica Enterprise credentials are provided in `.user` and `.password` configuration fields.

These credentials are stored in a configuration file along with the service installed on a Spotfire Server node, with access limited to administrators of this server.

## Data Function Execution

When the Spotfire Consumer user logs in to a browser and opens a Spotfire Analysis that is linked to a Statistica Workspace stored in the Enterprise database, Statistica data functions might be executed in Spotfire Server environment. This requires two logins.

- Spotfire Consumer log in to Spotfire Server
- Statistica user log in to retrieve the Workspace

When a data function executes, the current Spotfire user account is used by the Statistica Service to log in to Statistica.

> **Note:** When troubleshooting, set the log level in `custom.properties` to DEBUG and search for `Got spotfireUserName` to see what username was provided to the Statistica Service by the Spotfire Server.

When Spotfire and Statistica are configured to use Windows domain integrated login, then the Active Directory is responsible for Authentication. As the Spotfire Consumer has already logged in to the domain, there is certainty that *domain\fred* in Spotfire is the same *domain\fred* in Statistica.

For custom database users in Spotfire and Statistica user database, there is no certainty that *SPOTFIRE\fred* is the same person as *fred* within Statistica. If this scenario is required, then a policy needs to be established about creating users and using the same names within both systems. And the Statistica Service property would need to be configured as:

```
statistica.enterprise.spotfiredomain: SPOTFIRE
```

While SPOTFIRE is the default pseudo-domain, discuss this with your Spotfire administrator to confirm that this is true.

## Prohibited User Mappings

The `statistica.enterprise.prohibitedusermappings` property specifies custom database users that must not be mapped from Spotfire to Statistica. By default this is “admin;system”. A Spotfire administrator does not automatically get administrator rights within Statistica. To allow one or both of them, uncomment the property and remove the desired usernames, and add specific names you want to prohibit.

## Fallback User

The `statistica.enterprise.fallbackuser` property applies to the execution of a Data Function. By default this property is commented out and disabled.

This property can be used when you do not want employees to wait for the help desk to fix their login account within Statistica. It is important that DXP files always execute. All the Spotfire Consumers are trusted to see similar inputs and outputs.

This means that the user configuring the software can create a local user with Statistica Enterprise Manager. You can add this local user to this property and no other user must be

created within Statistica Enterprise Manager. Also, this local user needs read permission (authorization) on the Statistica Workspaces in Statistica Enterprise Manager.

Use of `fallbackuser` property in the following three cases:

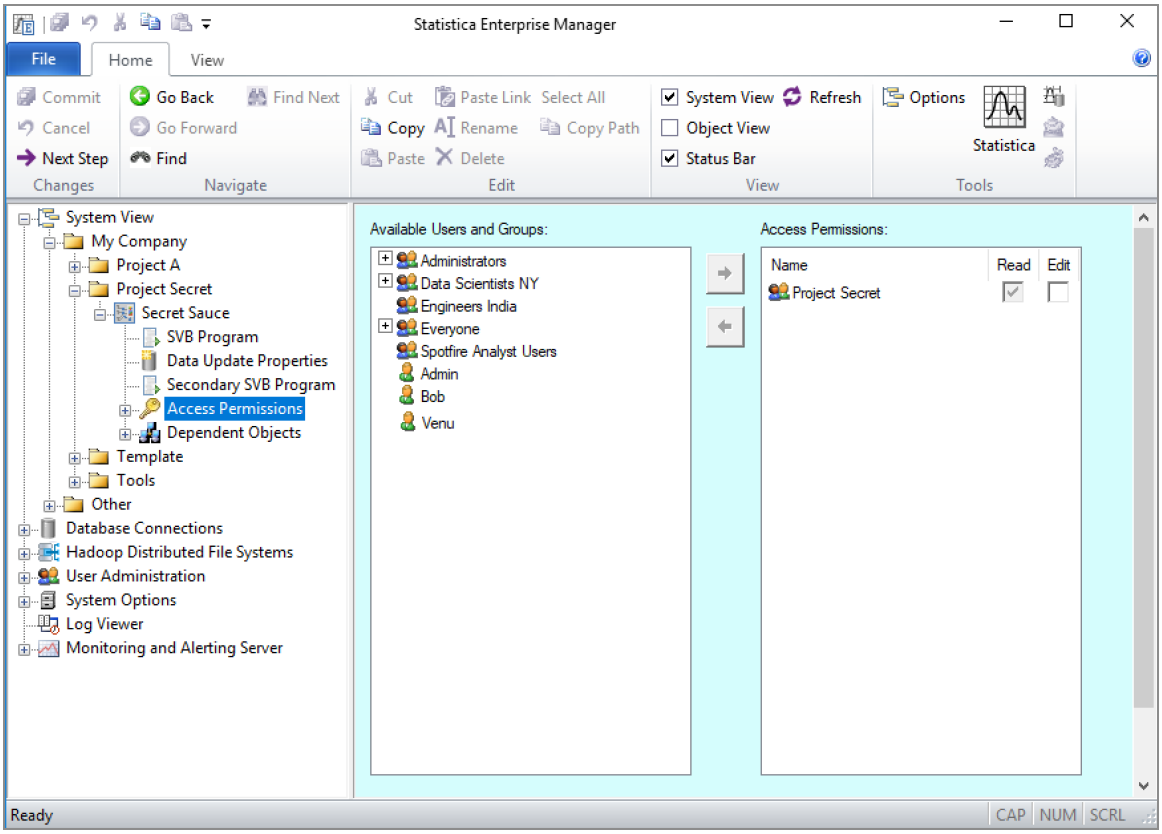
- Spotfire permission handles 100% authorization. User has the right to run a DXP file (Spotfire Analysis) and can also run the SDM file (Statistica Workspace) connected to it. In this case, you can create just one local user, and add it to this property.
- To track who executed the Statistica Workspace that is connected to the DXP file, you must use the audit trail within Statistica. In such cases, do not use this property.
- It takes days or a week to onboard employees and to get the new employee login created. You want new employees to be able to run DXP files if they have a Spotfire login, in this case this fall back user can help cover this timing issue.

# Product Connectivity - Authorization

Raw data that is in a file, `.csv`, needs to be authorized by policies to obtain access. For example, the company might require that data files are never emailed and must be shared using Google Drive or similar method which does have authorization by permissions.

Spotfire Statistica® Server and Spotfire® Data Science - Operations products provide the following authorization mechanism.

Metadata for objects like users, groups and their permissions (authorization), Files (Excel or Statistica Spreadsheet datasets), Data Configurations (SQL), and Analytic Configurations (workspaces) as mentioned earlier are managed using the Statistica Enterprise Manager application. These groups grant users edit and read (execute) permission for these objects. These are known as object groups.




In this image, the group **Project Secret** has read permission to the Workspace named **Secret Sauce**. Read permission grants read and execution permission. This Workspace can

be executed within the Statistica environment or within the Spotfire environment. It can be executed by a person logging in to a browser and an application making an API call.

Groups that are created for authorization tend to be named after teams, departments, or projects. These are object groups. They do not have any system permissions in them. These groups are just a list of people. For example:

- Data Scientists NY
- Project Secret
- Team North

# Spotfire® Service for Statistica Authorization

Spotfire consumers, who open a DXP file connected to an SDM file, require read permission on the SDM file. Read permission within the Statistica Enterprise Manager application also grants execution permission.

# User Interactions with Authentication - Authorization

A Statistica user has access to two workflow user experiences for retrieving input and generating output. The variety of input and output sources means that there may be security considerations for different workflows.

The users need this access to create analytic projects. You can create a project once and use it many times. The following list describes the two workflows.

- **Interactive:** User starts Statistica and opens a file or retrieves data from a database into a file. The user interacts (sees) the raw data. The user then selects algorithms off a menu to generate the analytics result.
- **Workspace (project):** User creates a Workspace by adding nodes (steps) that are connected to other nodes with arrows. This is a visual method to link the different steps required to complete a project; retrieve data, merge data, remove duplicates, basic statistics, neural networks to generate a model.

If the user owns one of the following products within Spotfire Data Science, there are additional user experiences.

- **Spotfire Statistica Server:** The user logs in to WebStatistica to execute a Workspace, select filters, and review results. Or the Workspace can be scheduled to execute at 10:00 PM every night and email the results to a manager.
- **Spotfire Data Entry Server:** The user logs in to manually enter data, validate data, and approve it. After entering the data, the user can execute a Workspace and review results.
- **Spotfire Monitoring & Alerting Server:** The user logs in to WebStatistica and reviews red/yellow/green alarms or receives an email alert.
- **Spotfire Live Score Server:** This is an API called to execute a Workspace or PMML model.
- **Spotfire Spotfire Analyst or Spotfire Spotfire Consumer:** The user opens Spotfire DXP file which executes a Statistica Workspace and displays results on the Spotfire UI.

# Restricting Access to Sensitive Data

To understand how data flows into the system, you can start with the Statistica client application. It can import and display data for a variety of file formats.

You can control access to this type of raw data by using file access permissions. For example, an admin grants a team access to read access for files in `\\file-server\teamA` directory. Or a company might require data files to be uploaded and shared using Google Drive or similar system.

The Statistica client application has two utility applications, Statistica Query and Advanced Query Builder, that can be used to retrieve data with a ADO.NET, ODBC or OLE DB database driver. The user can only retrieve data if they have a valid login and password for the database.

If you have installed Spotfire® Data Science - Operations or Spotfire Statistica® Server, explore the following questions:

- Since the application controls Authentication and Authorization, does company policy allow creation of system database login accounts? In other words, is it acceptable for the Database Connection to log in to Database "xyz" with a login "statistica"? And the groups configured within Statistica Enterprise Manager (Windows Domain) will grant access.
- What is the review process for SQL in the Data Configuration object? Should the Data Configuration call a stored procedure rather than SQL that was written within the application? If the raw data is not especially sensitive nature of data, then this question might be moot.
- Is there sensitive data embedded within a Workspace (Analysis Configuration) or File? If yes, you might need separate groups created for Authentication and Authorization.

The following data sources are supported; each may need a separate security group:

- A Statistica Workspace can call other Statistica Workspaces. Distinct workspaces may need distinct groups.
- Spotfire® Data Virtualization; publish an ODBC source
- Spotfire Spotfire® to access data sources using a data connection managed within the Spotfire Library

- Spotfire® Data Science - Team Studio in Hadoop or database; Statistica workspace calls Team Studio workflow and returns results into Statistica
- API - If the data source has an API, then the customer can write R, C#, Python, or Spark to access the data
- Hierarchical Process Cubes like Enterprise Resource Planning (ERP) such as SAP (MDX code)
- Data historian repositories such as the PI Data Historian from OSISoft, Inc. and the newer PI AF / EF
- Statistica spreadsheet has ODBC driver and can be queried with SQL. A Statistica spreadsheet with upper and lower spec limits for a QC chart might be stored with Statistica Enterprise Manager. Then the data is queried using a Query Spreadsheet node within a Workspace (Analysis Configuration).
- Amazon S3
- H2O
- Unstructured text (text mining)
- Manual data entry using Spotfire Statistica® Data Entry Server (double blind data entry with analytics)

# Considerations for Output

To understand how data flows out the system, remember that results can be saved into file formats listed previously. A Workspace can be configured to email analytic results to an end user.

In the Workflow section, different methods of reviewing analytic results in a web server are listed. Access to these are controlled with permissions.

A Workspace can be configured to write the results to a database. Specifically the Write Spreadsheet to Database node in a Workspace contains a database login/password. This node can insert a table, drop a table, delete all rows from a table, and then insert new data.

# Post-Installation Activities

Options for Statistica desktop interface and documents are stored in the `StatOpts.xml` file. For more information on the post-installation configuration, see the Spotfire Statistica® Options Configuration guide.

# Spotfire Documentation and Support Services

For information about this product, you can read the documentation, contact Spotfire Support, and join Spotfire Community.

## How to Access Spotfire Documentation

Documentation for Spotfire products is available on the Product Documentation website, mainly in HTML and PDF formats.

The Product Documentation website is updated frequently and is more current than any other documentation included with the product.

## Product-Specific Documentation

The documentation for this product is available on Spotfire Statistica® Product Documentation page and Spotfire Statistica®.

## How to Contact Support for Spotfire Products

You can contact the Support team in the following ways:

- To access the Support Knowledge Base and getting personalized content about products you are interested in, visit our product Support website.
- To create a Support case, you must have a valid maintenance or support contract with a Cloud Software Group entity. You also need a username and password to log in to the product Support website. If you do not have a username, you can request one by clicking **Register** on the website.

## How to Join Spotfire Community

Spotfire Community is the official channel for Spotfire customers, partners, and employee subject matter experts to share and access their collective experience. Spotfire Community offers access to Q&A forums, product wikis, and best practices. It also offers access to

extensions, adapters, solution accelerators, and tools that extend and enable customers to gain full value from Spotfire products. In addition, users can submit and vote on feature requests from within the Spotfire Ideas Portal. For a free registration, go to Spotfire Community.

# Legal and Third-Party Notices

SOME CLOUD SOFTWARE GROUP, INC. (“CLOUD SG”) SOFTWARE AND CLOUD SERVICES EMBED, BUNDLE, OR OTHERWISE INCLUDE OTHER SOFTWARE, INCLUDING OTHER CLOUD SG SOFTWARE (COLLECTIVELY, “INCLUDED SOFTWARE”). USE OF INCLUDED SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED CLOUD SG SOFTWARE AND/OR CLOUD SERVICES. THE INCLUDED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER CLOUD SG SOFTWARE AND/OR CLOUD SERVICES OR FOR ANY OTHER PURPOSE.

USE OF CLOUD SG SOFTWARE AND CLOUD SERVICES IS SUBJECT TO THE TERMS AND CONDITIONS OF AN AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER AGREEMENT WHICH IS DISPLAYED WHEN ACCESSING, DOWNLOADING, OR INSTALLING THE SOFTWARE OR CLOUD SERVICES (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH LICENSE AGREEMENT OR CLICKWRAP END USER AGREEMENT, THE LICENSE(S) LOCATED IN THE “LICENSE” FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE SAME TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.



Statistica, Spotfire, Process Tree Viewer, Process Data Explorer, Predictive Claims Flow, Live Score, Electronic Statistics Textbook, and Data Health Check, are either registered trademarks or trademarks of Cloud Software Group, Inc. in the United States and/or other countries.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

This document includes fonts that are licensed under the SIL Open Font License, Version 1.1, which is available at: https://scripts.sil.org/OFL

Copyright (c) Paul D. Hunt, with Reserved Font Name Source Sans Pro and Source Code Pro.

Cloud SG software may be available on multiple operating systems. However, not all operating system platforms for a specific software version are released at the same time. See the “readme” file for the availability of a specific version of Cloud SG software on a specific operating system platform.

THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. CLOUD SG MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S), THE PROGRAM(S), AND/OR THE SERVICES DESCRIBED IN THIS DOCUMENT AT ANY TIME WITHOUT NOTICE.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "README" FILES.

This and other products of Cloud SG may be covered by registered patents. For details, please refer to the Virtual Patent Marking document located at https://www.tibco.com/patents.