TIBCO® Data Science - Team Studio

Security Guidelines

Version 7.1.0 | September 2023

Contents

Environment Overview

Understanding the components, and the communication between the components of the TIBCO® Data Science - Team Studio environment is key to understanding how to build a more secure environment.

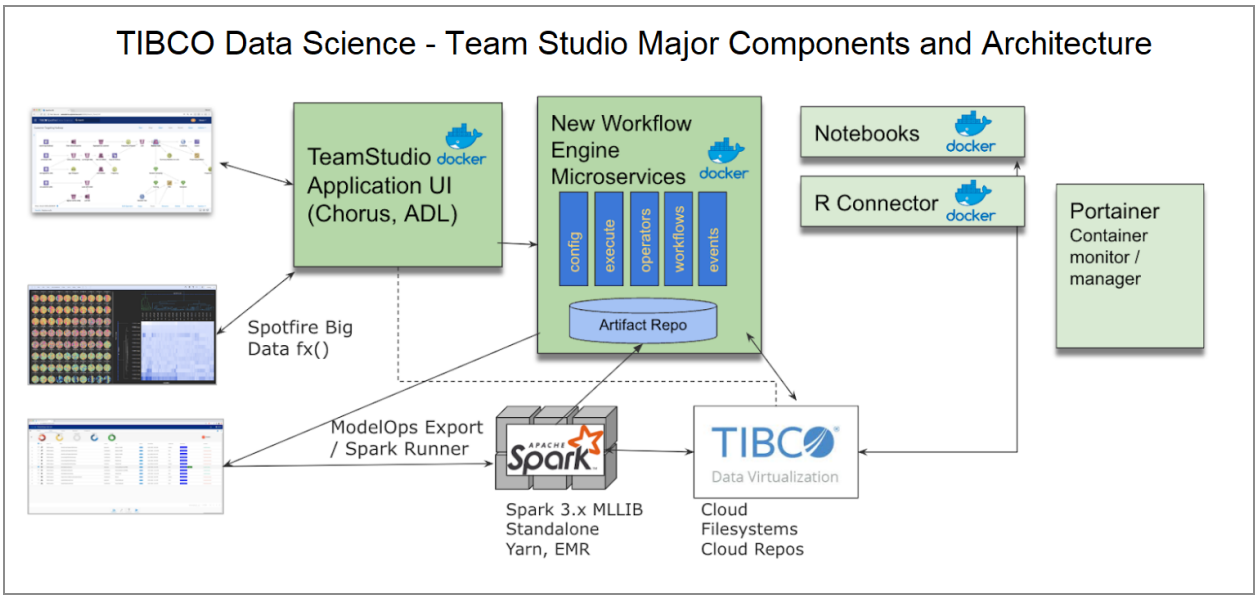
TIBCO Data Science - Team Studio is a collaborative, web-based user interface that you as a data scientists, data engineers, and business users use to create data science and data pipeline workflows with all sizes and velocities of data. The drag-and-drop visual workflows can be built with a minimum code. You can also use coding languages through Python notebooks to use the power of big data platforms, streaming sources, and more. TIBCO Data Science - Team Studio supports the entire data lifecycle, including data discovery, exploration, modeling, and deployment.

Most customers using TIBCO Data Science - Team Studio typically work with very large data and are from the high-tech manufacturing industry.

Product Connectivity

The following diagram shows all the components, as well as how data flows and network protocols are used in a typical TIBCO Data Science - Team Studio environment.

Figure 1: TIBCO Data Science - Team Studio Major Components and Architecture



TIBCO Data Science - Team Studio have the following components that are containerized using the Docker:

1. Application UI
    a. Chorus
    b. Alpine
    c. Notebooks
    d. R Connector
2. New Workflow Engine Microservices
    a. Configuration service (tds-configs)
    b. Modern workflows service (tds-workflows)
    c. Events service (tds-events)

    d. Execution service (tds-executions)
    e. Operator service (tds-operators)
    f. Resposilite
3. Nginx as reverse proxy

Ports and Protocols

You can use the following ports, connections, and protocols to secure TIBCO Data Science - Team Studio.

Public-Facing Client Connection Ports

The following table provides the public-facing client connection ports for the TIBCO Data Science - Team Studio:

Name	Default Port and Protocol	Function	Secure
Public HTTP port	80/TCP, if enabled	Non-secure communication with installed clients and web clients.	No
Public HTTPS port	443/TCP, if enabled	Secure communication with installed clients and web clients.	Yes

Outbound Connections

TIBCO Data Science - Team Studio is connected with the following outbound connections:

Connection	Definition
Data Sources	It is used to view, read, or write data to or from the configured Data sources in TIBCO Data Science - Team Studio.

Connection	Definition
Spark Cluster platforms	It is used to submit the workflow execution request to the configured Spark clusters.
LDAP (Lightweight Directory Access Protocol)	It is used to manage and authenticate users.

Authentication and Authorization

By default, TIBCO Data Science - Team Studio manages users through the PostgreSQL database. However, it can be configured to authenticate against an external LDAP (Lightweight Directory Access Protocol) server. The TIBCO Data Science - Team Studio collaboration framework uses the LDAPv3 server, including Active Directory support to manage and authenticate users.

Authentication

TIBCO Data Science - Team Studio have the following three methods to log in to Chorus for creating a valid session or obtaining an authentication token.

Methods	Definition
Internal Chorus authentication system	At the core of this component, Chorus stores the hash of a password with a unique salt on the user's table in the Chorus PostgreSQL database.
LDAP (Lightweight Directory Access Protocol)	The password is stored in an external LDAP system and the LdapClient is used to authenticate the password.
SAML (Security Assertion Markup Language)	An external SAML identity provider (IdP) is used to log in.

Authentication with HTTP Request

After logging in, Chorus verifies the authenticity of HTTP requests by the following methods:

1. Authentication via session
2. Authentication via token

Authorization

The roles assigned to the user authorize them to perform various tasks. For more information, see User Roles.

User Roles

User roles describe a person's expertise and role in the project. Each role enables a set of licenses that correspond to an administration or application role. This helps with team communication and messaging of what each person works on. You can have only one application and one administration role at a time. The available roles are:

Application Role

Determines the user permissions to work with the analytic workflows and data. The total number of application roles depends on the application license. The available application roles are:

Role	Definition
Analytics Developer	Analyzes the sets of data and finds common patterns, themes, and trends using the latest machine-learning techniques. The analytics developer focus on developing and implementing analytical solutions. They have the broadest permission. Analytical developers collaborate closely with data analysts and business users to comprehend their needs and transform them into practical analytical solutions.
Data Analyst	Examine large sets of data to extract meaningful insights and make data-driven recommendations. They employ various analytical tools to carry out tasks like data cleaning, data exploration, statistical analysis, and data visualization. Data analysts collaborate closely with business users to comprehend their goals and offer practical information to support decision-

Role	Definition
	making.
Collaborator	Refers to a team member who actively participates in the data analysis process. Individuals from several departments or job roles who provide domain-specific knowledge or skills are considered collaborators. They provide input on data requirements, offer insights and perspectives, and work with data analysts to ensure that the analysis aligns with the organization's goals.
Business User	A professional who applies data and analytics to their roles to help them make informed decisions. Business users rely on the insights provided by data analysts and analytical developers to comprehend market trends, customer behavior, and operational performance. These roles often collaborate with one another and work together to use data for better business outcomes. They have the fewest permissions.

Administration Role

Determines the user permissions to manage the items in the application. You can have only one administration role at a time. The available administration roles are:

Role	Definition
Application Administrator	Responsible for managing and maintaining the application. Their primary focus is on the operational aspects of applications, ensuring their availability, performance, and security. The number of application administrators depends on the licensing limits.
Data Administrator	Manages the data source associations for workspaces, ensuring data quality, security, and compliance. There are no limits on the number of data administrators.

Data Access Control

You must have the Administrator or Data Administrator credentials to control the data access.

Data Visibility and Control

The data visibility in TIBCO Data Science - Team Studio is a system for managing the data sources that users can see and access within workspaces. The goal is to provide more granular control and security by allowing users to know about certain data in the application.

The TIBCO Data Science - Team Studio data visibility system has four tenets. All of them come together to form a robust and cohesive data visibility offering.

The data sources can be global, that is, they can be designated as "public" or they can be scoped as "limited", restricting their visibility and available to only the workspaces with which they are associated. For more information on data source visibility, see "Controlling Data Source Visibility" in *TIBCO® Data Science - Team Studio Installation and Administration*.

Data Access Permissions

You can change the level of permissions on the data source could be changed. A user might be able to see a data source in the Data Sources section in their workspace, but they cannot access it until they have permission. For more information on data access permissions, see the "Controlling Data Source Permissions" topic in *TIBCO® Data Science - Team Studio Installation and Administration*.

For Monitoring the Logs

TIBCO Data Science - Team Studio provides a feature to monitor the logs of all microservices with the help of monitoring tools. The user must install the docker monitoring tools such as Portainer, DataDog, or Prometheus to access the container logs. The user can also access the container logs from the command line using the following docker commands.

```
docker logs -f <container_name>
```

The installer is equipped to install a free version of Portainer. If the system administrator opted to install this during the configuration stage, you can access the monitoring tool by appending `/monitoring/` at the end of the TIBCO Data Science - Team Studio URL.

The monitoring tool helps in tracking the issues related to modern workflow. Along with the Spark cluster logs, you can also check the logs of *tds-workflows* and *tds-executions* microservers.

Session Management

A session is created whenever a user accesses the TIBCO Data Science - Team Studio Server from a web browser. A session ID is valid across the TIBCO Data Science - Team Studio Server environment.

Session IDs

The TIBCO Data Science - Team Studio session IDs are sequences of randomly generated 16-bytes strings. The JSESSIONID is the HTTP cookie by which the session IDs are propagated through the layers of the TIBCO Data Science - Team Studio architecture. All information associated with the session is stored on the server-side.

Session Rotation

The session IDs in TIBCO Data Science - Team Studio are rotated (replaced) when a user authenticates and change their own passwords in the TIBCO Data Science - Team Studio Server database. The changes in credentials in the external authentication systems do not have any impact on the active sessions. When a session ends, the session is invalidated (deleted) on the Server side. The session IDs may remain in a client cookie store but it does not refer to any active session. Thus, any subsequent attempts to use the IDs are ignored.

Session Timeouts and Configuration

You can configure the aspects of the session management in TIBCO Data Science - Team Studio. For example, the user session timeout.

The login session expires after a configurable amount of time set in the *chorus.properties* files. The default time is 480 minutes or 8 hours.

Cryptography

The Administrator can configure the authentication data and cryptographic keys for user-facing services. The datasource passwords are encrypted before being stored in the TIBCO Data Science - Team Studio internal database. The encryption algorithm is Advanced Encryption Standard (AES) with 128-CBC. A secret key is used for verifying the integrity of signed cookies, which can be regularly rotated (replaced) by the system administrator.

Data at Rest

Data at rest is data stored, either temporarily or permanently. Data at rest has certain encryption types, or no encryption, depending on where it is being stored.

- Data stored in the TIBCO Data Science - Team Studio database is not encrypted, except for especially sensitive data like passwords for data source configurations.
- The temporary files stored on the Alpine server file system are not encrypted.

Data in Motion

Data in motion has certain encryption protection, depending on how and where it is moving in the TIBCO Data Science - Team Studio environment.

- Communication between the TIBCO Data Science - Team Studio Server and external interfaces (UI/API calls) is always encrypted using Transport Layer Security (TLS).
- Communication with the external data sources can be secured by either TLS or vendor-specific encryption protocols.

TIBCO Documentation and Support Services

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

How to Access TIBCO Documentation

Documentation for TIBCO 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 following documentation for TIBCO® Data Science - Team Studio is available on the TIBCO® Data Science - Team Studio Product Documentation page:

- *TIBCO® Data Science Team Studio Release Notes*
- *TIBCO® Data Science Team Studio Installation and Administration*
- *TIBCO® Data Science Team Studio User Guide*
- *TIBCO® Data Science Team Studio Security Guidelines*
- *TIBCO® Data Science Team Studio API Documentation*
- *TIBCO® Data Science Team Studio Web Help*

How to Contact Support for TIBCO 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 our product Support website. If you do not have a username, you can

request one by clicking **Register** on the website.

How to Join TIBCO Community

TIBCO Community is the official channel for TIBCO customers, partners, and employee subject matter experts to share and access their collective experience. TIBCO 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 TIBCO products. In addition, users can submit and vote on feature requests from within the TIBCO Ideas Portal. For a free registration, go to TIBCO 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.

TIBCO, the TIBCO logo, the TIBCO O logo, Spotfire, and Alpine Data Labs 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.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only. You acknowledge that all rights to these third party marks are the exclusive property of their respective owners. Please refer to Cloud SG’s Third Party Trademark Notices (https://www.cloud.com/legal) for more information.

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.