WHITEPAPER

Overview:
Platform Security

RAPIDMINER

RAPIDMINER

Table of Contents

Overview	**3**
Layered Security for the Enterprise	**4**
Security in the RapidMiner Platform	**4**
Verifying User Identities	**5**
Controlling User Access	**6**
Protecting Company Data	**7**
Tracking & Recording Data	**8**
Wrapping Up	**9**
Key Feature Overview	**10**

Before making full use of your company's data, you need to establish a secure process for working with it. Even the most well-planned machine learning projects can fall flat without protocols to manage user access and track your data over time.

The tools that you choose have a strong impact here. Any technology that uses your organization's data needs to comply with the protocols you've established to ensure security.

In this overview, we'll take you through the security infrastructure of RapidMiner's platform—infrastructure that'll allow you to spend less time worrying about compliance and more time gathering actionable insights from your data.

Overview

Most forward-thinking organizations know that there's a tremendous amount of value that can be extracted from their data. Today, this goes beyond reporting on a company's current state & performance against key metrics. When implemented correctly, a strong data science practice can enable your organization to gain insights on customer behavior, assess risk, and predict the likely outcomes of business decisions.

As more companies realize that data is the new oil, the rate at which they gather & share information is growing. Companies are increasingly reliant on the digital exchange of information—the faster you can share knowledge, the better chance you have at driving desirable business outcomes.

"Not only has the sheer volume of available data—mainly driven by the IoT—grown exponentially over the past five years... but new tools [have] been developed for turning this flood of raw data into insights and eventually into action."

- ACHIEVING BUSINESS IMPACT WITH DATA,
MCKINSEY & COMPANY

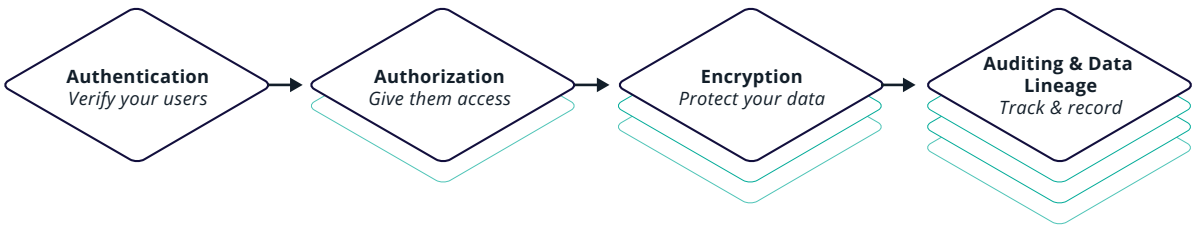
With that opportunity comes risk. While organizations focus on extracting value from their data, malicious employees and criminal elements focus on stealing and misusing it. In recent years, several high-profile data breaches have put customer data at risk, incurred a great deal of cost, and damaged the brands of the organizations who've suffered them.

Before fully embracing the competitive advantage that data science can provide, it's equally important to understand—and proactively address—the risk associated with working through large datasets. In the following sections, we'll walk through the security infrastructure of RapidMiner's platform and show that you don't need to trade security for actionable insight.

Layered Security for the Enterprise

Before exploring security measures specific to RapidMiner, it's important to note that those measures are built on a 4-layer data security implementation model.

Data Security Implementation Model



Developing a secure process for working with your data requires a focus in two key areas: having full control over who is accessing your data, and visibility into the data itself. When both areas are addressed, you empower the right people to make strategic decisions—and know that the information behind those decisions is credible.

Security in the RapidMiner Platform

RapidMiner's security protocols give your account administrators & IT staff the ability to verify users' identities, set their permissions, encrypt your company's data, and track it over time. In the following sections, we'll outline the scope & overall impact of these key areas and share the RapidMiner security features associated with each.

- **Verifying User Identities**
- **Controlling User Access**
- **Protecting Company Data**
- **Tracking & Recording Data**

Verifying User Identities

This particular security protocol refers to user authentication, or the measures put in place to ensure that users are who they claim to be. Authentication methods such as single sign-on (SSO) and SAML have become increasingly common as organizations adapt to employees' desire to access their work apps from multiple Internet-connected devices.

Authentication in RapidMiner

In order to ensure login security in RapidMiner AI Hub, we've incorporated Keycloak as the platform's primary authentication layer. This provides a single sign-on experience for all components of the RapidMiner platform, while using standard protocols for authentication and authorization. Admins have the ability to set password complexity requirements and configure two-factor authentication.

Additionally, Keycloak enables identity & user federation using standards such as OpenID Connect, SAMLv2.0, and LDAP—providing an SSO experience for users across all enterprise applications.

RapidMiner Security Flow

Key Authentication Features in RapidMiner

- Single-Sign On/Sign Out (SSO)
- Social Login (Google, GitHub etc.)
- Two-Factor Authentication (2FA)
- SAML Authentication Support
- OAuth 2.0 Support
- OpenID Connect (OIDC)

Controlling User Access

Data Access Security *allows administrators to determine what types of data users can see (and work with) on their networks. This process doesn't just improve security—it allows workers to focus on what they need to do their jobs & not be distracted by the rest.*

Authorization in RapidMiner

Admins can establish role-based authorization to separate functionality that's meant for admins vs. other users, and create custom role mappings to restrict users to the parts of the platform they'll be working in.

By integrating Keycloak, RapidMiner also provides the ability to create permissions, associate them with authorization policies, and enforce those authorization decisions within the platform. We've implemented a file-level permissioning scheme so admins can control access to data stored in RapidMiner's central repository—to take this a step further, AI Hub also allows you to protect access to files that are being used in open & collaborative projects.

Lastly, on big data—RapidMiner respects whatever security policies you've established in your Hadoop cluster, down to row and column-level filtering that you've defined for users.

Key Authorization Features in RapidMiner

- OAuth 2.0 Support
- OpenID Connect Support
- Built-in Admin Console
- Token Mappers
- Hadoop Security Support
- Groups & Role Mappings

In production environments, RapidMiner administrators can assign special permissions that allow individual users to execute or schedule jobs.

Protecting Company Data

From a security standpoint, encryption is one of the most crucial steps an organization can take to protect its data. Data encryption hides data so that only people with a password or decryption key can view it. Encryption algorithms impact multiple areas of overall security including authentication and data integrity.

Encryption in RapidMiner

RapidMiner AI Hub supports the Transport Layer Security (TLS) protocol to encrypt data as it travels over your network. Connection metadata is also encrypted, allowing access to business critical systems without exposing credentials to end users or third parties.

For organizations that choose to enable LDAP authentication, AI Hub also allows for the encryption of LDAP properties to better protect your configuration.

Lastly, it's worth noting that repositories as a whole are not encrypted on the application level—RapidMiner relies on file system-level encryption.

Key Data Protection Features in RapidMiner

- Transport Layer Security (TLS) Support
- Encryption of LDAP Properties
- Hadoop HDFS Encryption

Tracking & Recording Data

An administrator's ability to monitor and audit data access no matter where it resides (ex. database, server, Hadoop cluster) is a key component in ensuring a secure data science practice. Understanding data lineage is crucial here—in order to trust the data being used to make decisions, an organization must first understand its' origins and be able to map its' transformation over time.

Admin Console

Auto Logging

Rollback Capability

Auditing & Data Lineage in RapidMiner

In order to give administrators the ability to audit processes and models over time, RapidMiner provides logging and versioning capability. Logs are generated across the entire product suite, allowing admins to see which users are logging in, and when. All authentication and authorization-related changes in RapidMiner AI Hub are tracked in an audit log.

AI Hub is also built to help you track data lineage. In a given process, metadata can show you where a model originated as well as how a piece of data has transformed. Projects also have a commit history, which shows who has made modifications to groups of data & when.

Key Tracking & Recording Features in RapidMiner

- Versioning & Rollback Capability
- Automatic Logging
- Built-in Admin Console
- Event Mapping
- Admin Session Management
- Groups & Role Mappings

Wrapping up

Before your organization can reap the benefits of a strong data science practice, it's crucial to establish a secure process for working with your data. The strongest processes account for verifying users, controlling their level of access, protecting your data, and understanding its lineage.

RapidMiner allows enterprises to do just that. By giving your organization full control over who can access the platform and what they can see, RapidMiner's security protocols ensure that sensitive information can only be seen by parties who you've authorized.

In order to maintain the integrity of that information, the platform also gives users the ability to track changes that are being made to it over time.

With a secure process in place, employees within your organization can focus on gathering actionable business insights and using them to drive you forward.

RAPIDMINER

For those driven to accelerate the pace of transformation, RapidMiner is the enterprise-ready data science platform that amplifies the collective impact of your people, expertise, and data for break-through competitive advantage. RapidMiner's data science platform supports all analytics users across the full AI lifecycle. The RapidMiner Academy and Center of Excellence methodology ensure customers are successful, no matter their experience or resource levels. Since 2007, more than 1 million professionals and 40,000 organizations in over 150 countries have relied on RapidMiner to bring data science closer to their business.

Key Feature Overview

Authentication

- Single-Sign On/Sign Out (SSO)
- Social Login (Google, GitHub etc.)
- Two-Factor Authentication (2FA)
- SAML Authentication Support
- OpenID Connect (OIDC)

Authorization

- OAuth 2.0 Support
- OpenID Connect Support
- Built-in Admin Console
- Token Mappers
- Admin Session Management
- Groups & Role Mappings

Data Protection

- Transport Layer Security (TLS) Support
- Encryption of LDAP Properties
- Hadoop HDFS Encryption

Auditing & Data Lineage

- Versioning & Rollback Capability
- Automatic Logging
- Built-in Admin Console
- Event Mapping
- Admin Session Management (for data tracking & rollback)
- Groups & Role Mappings