

Security Overview

in the RapidMiner Platform

Secure IT infrastructure and systems are a must for organizations. Security is an integral part of all internal processes as organizations seek to lock down data and meet security and compliance standards. Predictive analytics does not operate outside these initiatives and must be brought into the security fold.

In this whitepaper, we discuss how the RapidMiner Platform complies with security standards – providing authentication, authorization, accountability, and data encryption.

Supports Security Standards Including:

- SSL
- Active Directory
- Kerberos
- Apache Ranger
- Apache Sentry

The Challenge

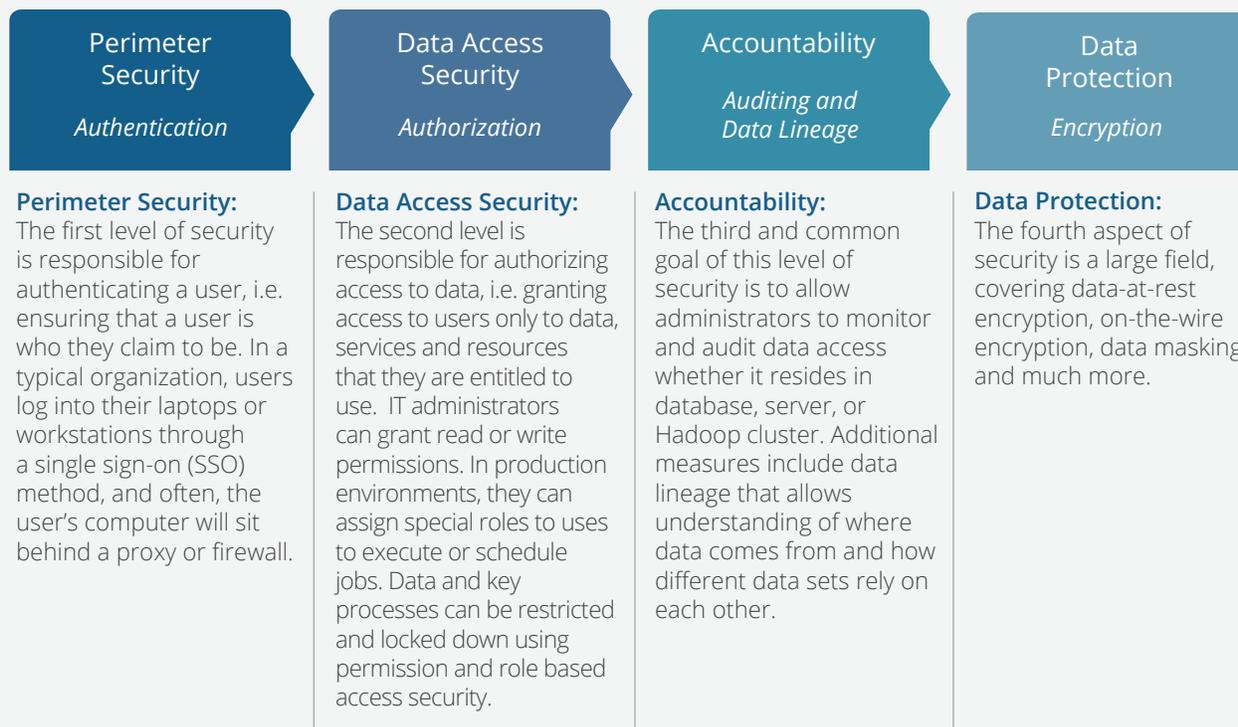
Data has become ubiquitous in an organization and comes in many different formats. Flat files (i.e. CSV, XLSX, etc.) are everywhere, email between team members is growing, Database Administrators provide SQL data dumps to Data Scientists, and organizations capture customer transactional data at ever increasing velocities in Hadoop clusters. This data is untapped competitive advantage for an organization and simultaneously it's weakness - a Schrödinger's cat!

While organizations focus on extracting value from their data, malicious employees and criminal elements focus on stealing and misusing that data. Several high-profile data breaches have put customer data at risk, incurred a great deal of cost, and hurt the brand of the organization that suffered the data breach.

While some of security is driven by hardware driven and sound IT policy (i.e. disabling USB ports or not allowing email attachments), much of it needs to be handled at the data and application level.

DATA SECURITY IMPLEMENTATION MODEL

The standard security model is a 4-layer system:
Perimeter Security, Data Access Security, Accountability, and Data Protection.



Security in the RapidMiner Platform

RapidMiner is at the forefront of data science, bringing ease-of-use and visual analytics to data scientists who are looking to extract value from their data, regardless of where it lies.

For **Perimeter Security** needs, RapidMiner can sit easily behind a corporate firewall/proxy and all products can be silently installed by an IT administrator to control what version and extensions can be used by end users. RapidMiner Server and RapidMiner Radoop integrates LDAP or Active Directory to obtain user information, thereby satisfying your Perimeter Security needs. With Hadoop, this is usually solved with Kerberos, a well-known standard for implementing authentication. RapidMiner Radoop supports Kerberos and integrates with major Hadoop vendors for their perimeter security methods.

RAPIDMINER SECURITY FLOW

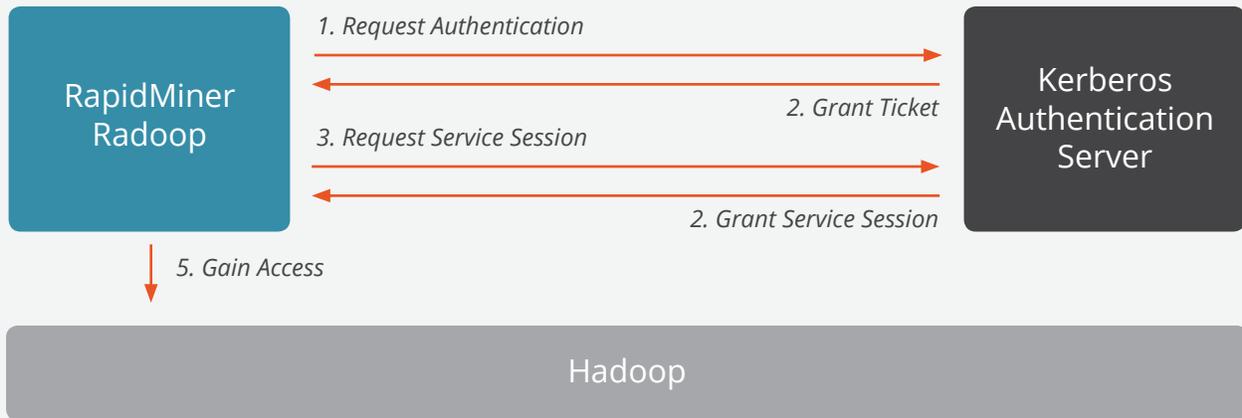


For **Data Access Security**, RapidMiner Server can “lock down” data and production processes via file permissions (i.e. Read, Write, Execute). Sometimes users are looking for more fine-grained authorization capabilities (e.g., column level or cell-level). RapidMiner Radoop integrates with Apache Sentry and Apache Ranger to deliver data access security inside your Hadoop cluster.

For **Accountability**, RapidMiner provides logging and versioning capability to support auditing of processes and models. Logs are generated across the entire product suite. Authorized users can generate new versioning points where any work in production, or non-production, can be rolled back to a previous point on RapidMiner Server, if needed. RapidMiner Radoop, supports Cloudera Navigator and Apache Ranger.

For **Data Encryption**: Data and metadata can be anonymized between internal endpoints using Obfuscation and De-obfuscation in RapidMiner Studio and RapidMiner Server. Additionally, RapidMiner Server supports HTTPS and HTTP with SSL which can be configured as part of installation. RapidMiner Radoop supports Hadoop HDFS encryption and wire encryption.

KERBEROS AUTHENTICATION



Summary

Compliance with Security Standards is not optional for enterprise software. Data, in many different formats is assaulting an organization daily with every growing volume. This data is an untapped competitive advantage for an organization but also its greatest weakness with respect to security. Data Scientists routinely use these disparate data sources to extract competitive advantage but use many different software in their toolkit to do so. While each separate piece of software might comply with one or two parts of the standard security implementation model, taken they might not work together at all.

The RapidMiner Platform was designed to comply with all aspects of the standard security implementation model and integrates easily into an organization's IT infrastructure. Whether in development or production, components of the RapidMiner Platform easily work together to provide security from start to finish.



10 Milk Street, 11th Floor
Boston, MA 02108

RapidMiner is a powerful & collaborative open source data science platform. RapidMiner offers more machine learning and data prep functions than any other visual solution, and has a community of over 200,000 users.

For more information, visit www.rapidminer.com

©2017 RapidMiner, Inc. All rights reserved