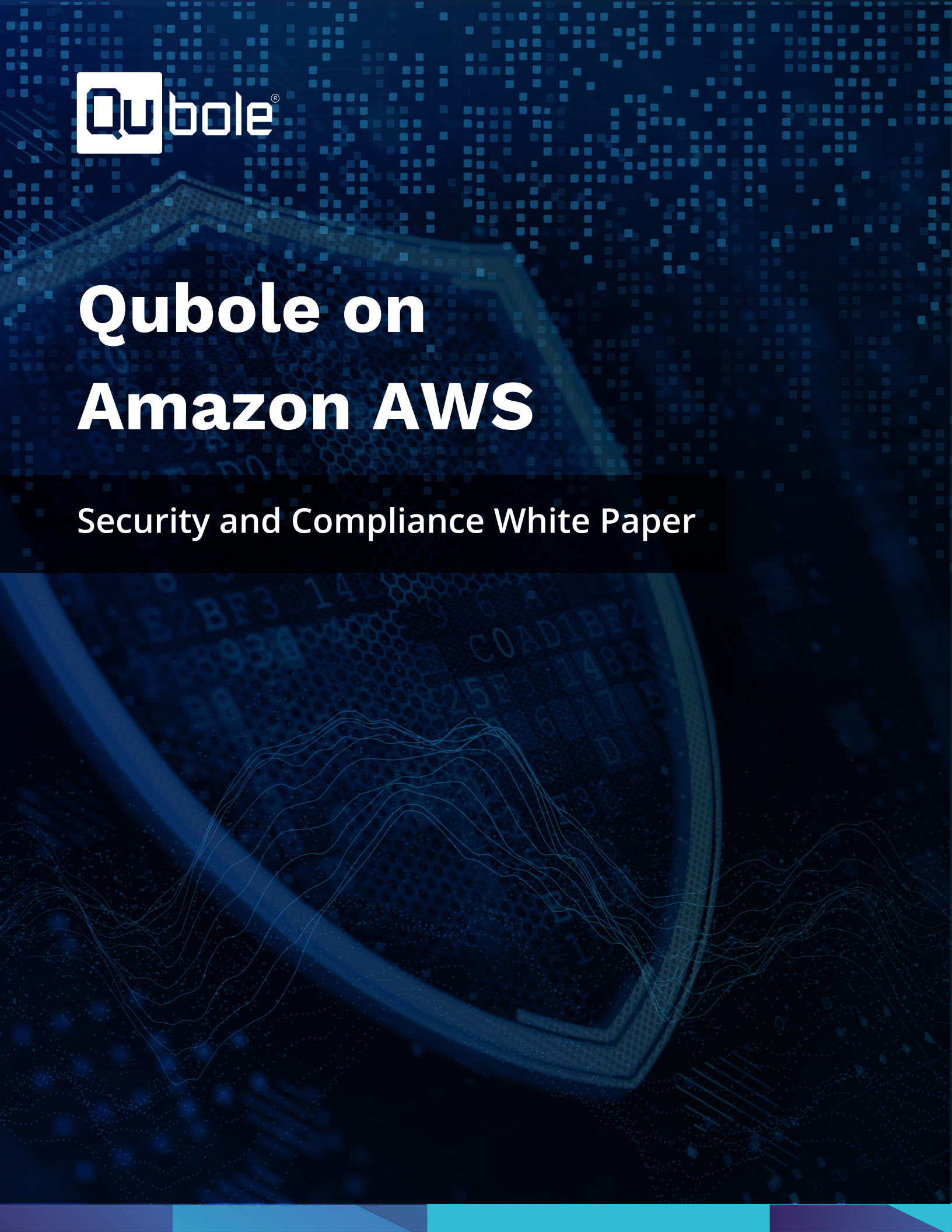


Qubole on Amazon AWS

Security and Compliance White Paper

Qubole
469 El Camino Real, Suite 201
Santa Clara, CA 95050
www.qubole.com
(855) 423-6674

Trademarks

Qubole is a registered trademark. All other trademarks or registered trademarks are the property of their respective owners.

Disclaimer

The information provided in this document is provided "as is" without warranty of any kind. Qubole disclaims all warranties, either express or implied, including the warranties of merchantability and fitness for a particular purpose. In no event shall Qubole be liable for any damages whatsoever including direct, indirect, incidental, consequential, loss of business profits or special damages, even if Qubole or its suppliers have been advised of the possibility of such damages.

Document Lifetime

Qubole may occasionally update online documentation between releases of the related software. Consequently, if this document was not downloaded recently, it may not contain the most up-to-date information. Please refer to **www.qubole.com** for the most current information.

Where to get help

Qubole support, product, and licensing information can be obtained as follows.
Product information — For documentation, release notes, software updates, or for information about Qubole products, licensing, and service, see the Qubole website at: **www.qubole.com**

Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Please send your opinion of this document to: **support@qubole.com**

If you have issues, comments, or questions about specific information or procedures, please include the title and, if available, the part number, the revision, the page numbers, and any other details that will help us locate the subject that you are addressing.

TABLE OF CONTENTS

TABLE OF CONTENTS

INTRODUCTION

As the leading open data lake platform, Qubole™ is passionate about making data easily accessible for ad-hoc analytics, streaming analytics and machine learning. Whether they're using Amazon Web Services, Google Compute, Microsoft Azure or Oracle Cloud, our customers process nearly an exabyte of data every month and we know that guarding customer data, enforcing proper security measures, regulatory compliance, and overall trust are essential to your success. This paper discusses the security strategies we use to protect your information and provides specific details of how our security model works with Amazon Web Services (AWS).

ORGANIZATIONAL SECURITY—OUR PHILOSOPHY

We believe that availability, security, governance and privacy are essential elements of the Qubole Open Data Lake Platform. As such, security is integral to our culture and our on-going mission. Security awareness at Qubole far exceeds the minimal standards required by auditors and compliance organizations and is incorporated in all aspects of our business model—from product design and development, to day-to-day corporate operations, we are constantly building and improving our robust organizational security programs.

Personnel Security

Qubole's personnel practices apply to all members of our workforce —regular employees and contractors— who have access to information systems. All Qubolers are required to understand and follow internal policies and standards. Before gaining initial access to systems, Qubolers must agree to confidentiality terms, pass a rigorous background screening, and attend annual security training. Upon termination, all access to our systems is immediately blocked.

Security And Privacy Training

During their tenure, all members of our workforce are required to complete a privacy and security training refresh at least annually. These courses cover privacy and security topics, including device security, acceptable use, preventing malware, physical security, data privacy, account management, and incident reporting. Qubolers are also required, on an annual basis, to acknowledge that they've read and will follow Qubole's information security policies.

Certain professionals, such as engineers, operators, and support personnel who may have elevated access to systems or data, receive additional job-specific training on privacy and security. Qubolers are also required to report security and privacy issues to appropriate internal teams and are informed that failure to comply with acknowledged policies may result in consequences up to, and including, termination.

QUBOLE'S SECURITY PROFESSIONALS

Qubole employs a Chief Security Officer (CSO) who functions as the organization's top executive responsible for security, compliance, and privacy. Our CSO serves as the company's business leader responsible for developing, implementing, and managing the organization's corporate security vision, strategy, and programs.

The CSO is supported by the other members of Qubole's Security Team, which currently consists of security professionals with more than 75-years of combined privacy experience. This team focuses on Product Security, Security Operations, Computer Security Incident Response, Risk, and Compliance. Together, they divide responsibilities for key aspects of our security program including:

Application Security

- Secure development practices and standards
- Ensure security risk assessments
- Design and code review services for detecting and removing common security flaws
- Developers are trained and understand secure coding practices and data privacy
- Developers understand risk issues

Product Security

- Design—Guide development by reviewing the design. This is the best stage to find a concern before it ever becomes code.
- Build—Code exists but it is in flux. Finding and fixing individual issues in the code through advice and awareness. The product security team serves as consultants to engineering, answering questions like "is this secure?" or "how should I write this?" and proactive outreach.
- Ship—A thorough review by the product security team, if appropriate before launch. Our product security team holds itself to an SLA and we work hard to keep launches on track.
- Live—Software spends most of its lifetime in this stage, so our security efforts don't stop at launch.

Security Operations

- Build and operate security-critical infrastructure including Qubole's public key infrastructure, event monitoring, and authentication services
- Maintain a secure archive of access privilege logs
- Consult with operations personnel to ensure the secure configuration and maintenance of Qubole's production environment

CSIRT

- Respond to alerts related to security events on Qubole systems
- Manage security incidents
- Acquire and analyse threat intelligence

Risk and Compliance

- Coordinate penetration testing
- Manage vulnerability scanning and remediation
- Coordinate regular risk assessments, and define and track risk treatment
- Manage the security awareness program
- Coordinate audits and maintain security certifications
- Respond to customer inquiries
- Review and qualify vendor security posture

All members of Qubole's Security Team are active participants in the larger information security community to improve the overall state of the art of information security and to maintain their own expertise.

Policies And Standards

Qubole maintains a set of policies, standards, procedures and guidelines ("security documents") that provide our workforce with the "rules of the road." Our security documents help ensure that Qubole customers can rely on us to behave ethically and for our service to operate securely. Security documents include, but are not limited to:

- Fair, ethical, and legal standards of business conduct and acceptable use of systems
- Classification, labelling, and handling rules for all types of information assets
- Practices for worker identification, authentication, and authorization for access to system data and secure development, acquisition, configuration, and maintenance of systems
- Workforce requirements for transitions, training, and compliance with ISMS policies
- Use of encryption including requirements for when and where to use it
- Description, schedule, and requirements for retention of security records and planning for business continuity and disaster recovery
- Classification and management of security incidents and change control
- Regular use of security assessments such as risk assessments, audits, and penetration tests

AUDITS, COMPLIANCE AND 3RD-PARTY ASSESSMENTS

Qubole evaluates the design and operation of its overall ISMS for compliance with internal and external standards. We engage credentialed assessors to perform external audits at least once per year including the following organizations.

Cloud Security Alliance

Qubole has completed the Consensus Assessment Initiative Questionnaire (CAIQ). This program outlines more than 200 questions relating to cloud services, provisioning, storage, security, availability, confidentiality, and privacy. A copy of this document is available from your account manager or account executive. The current Cloud Security Alliance CAIQ is free to download and can be found **here.**To learn more about Qubole's responses to the CAIQ, please reach out to Qubole Sales.

AICPA Service Organization Controls 2 Type 2 (SOC2)

Qubole has successfully completed the SOC 2 Type II examination. The report generated from this extensive evaluation was conducted by an American Institute of CPAs (AICPA) accredited firm that attests that Qubole meets stringent guidelines for data security.

SOC 2 reports are designed to provide information and assurances that the service delivery processes and controls used by a SaaS provider meet certain high standards and guidelines for data security, confidentiality and availability. There are two types of SOC 2 reports: Type I and Type II. We opted for the much more rigorous Type II report because security is of the utmost importance to us and to our many clients who entrust us to power their big data in the cloud initiatives. You can find our blog announcing this **here.**

Privacy Shield

Qubole has engaged with TrustArc (formerly TRUSTe) to complete and attest to compliance with the US Privacy Shield regulation around privacy and transfer of EU Personal Data to the United States.

GDPR and CCPA

The General Data Protection Regulation (GDPR) and the California Consumer Protection Act (CCPA) are new Data Privacy legislations to provide consumers the right to act on their personal data collected by 3rd parties. These regulations outline the importance of data subject's rights to the privacy of their personally identifiable information (PII). Qubole fully supports the GDPR and CCPA as an opportunity to deepen its commitment to data protection. Qubole also complies with the GDPR and CCPA in the delivery of our service to customers and is fully prepared to handle the intricacies of the areas related to GDPR and CCPA legislation. Qubole will also continue to enhance data protection and compliance in the following areas:

1. Accountability, policies, and procedures
2. Compliance and risk activities
3. Implementing additional security measures
4. Notification and reporting requirements for data breaches

Qubole has partnered with a 3rd party to continue to align innovation in big data to meet the exacting requirements of the GDPR. In addition, Qubole has created a Data Processing Addendum as an attachment to its Master Services Agreement. This document supports our commitment to this important legislation and is available through your sales representative.

HIPAA Compliance

Qubole has attested through a 3rd party auditor to being 45 CFR Part 164, Subpart C compliant. This includes Administrative, Technical and Physical safeguards as well as organizational requirements.

ISO-27001

Qubole believes that compliance must be built throughout the Open Data Lake Platform and, as such, our processes, procedures, controls, operations and activities align with the ISO-27001 standards and are reflected in our policies and other attestation and certification work. Qubole will certify ISO-27001 within the next twelve months.

Audit results are shared with senior management and all findings are tracked to resolution.

Penetration Testing

Qubole engages independent entities to conduct regular application and infrastructure-level penetration tests. Results of these tests are shared with senior management. Qubole's Security Team reviews and prioritizes the reported findings, creates tasks for engineering to resolve any deficiencies and tracks them to resolution. Qubole will share a summary of the most recently performed tests under non-disclosure upon request; please reach out to your Qubole account representative. Customers wishing to conduct their own penetration test of Qubole's system can contact their account representative regarding authorization to perform this test. Qubole aligns our penetration test against OWASP Top Ten. In summary, some of the areas covered are listed below:

A1. **Injection Flaws**—Qubole not only evaluates these flaws during the penetration test but also uses Static Code Analysis Testing during development.

A2. **Broken Authentication**—Qubole uses mechanisms to strengthen and securely store passwords, secrets and tokens.

A6. **Security Misconfiguration**—Qubole removes default components including default credentials. Qubole turns off and disables default services and hardens the baseline of compute systems.

A7. **Cross Site Scripting (XSS)**—Qubole evaluates the three primary types of XSS and has libraries in place in code to sanitize and prevent exploitation.

A10. **Insufficient Logging and Monitoring**—Qubole centrally logs all critical system and application logs. Qubole also has built-in anomaly detection for questionable events and actions and monitors for abuse and attacks through log analytics.

DESIGN SECURITY

Qubole assesses the security risk of each software development project according to our Secure Development Lifecycle. Before completion of the design phase, we undertake an assessment to qualify the security risk of the software changes introduced. This risk analysis leverages both the OWASP Top 10 and the experience of Qubole's Product Security team to categorize every project as High, Medium, or Low risk.

Based on this analysis, Qubole creates a set of requirements that must be met before the resulting change may be released to production. All code is checked into a version-controlled repository. Code changes are subject to peer review and continuous integration testing. Qubole's Security Team also operates continuous automated static analysis using advanced tools and techniques. Significant defects identified by this process are reviewed and followed to resolution by the Security Team.

PROTECTING CUSTOMER DATA

The focus of Qubole's security program is to prevent unauthorized access to customer data. To this end, our team of dedicated security practitioners, working in partnership with peers across all our teams, take exhaustive steps to identify and mitigate risks, implement best practices, and constantly evaluate ways to improve our security measures.

Data Encryption

Qubole supports end-to-end encryption throughout Qubole services as defined below. Encryption protects the confidentiality of customer data ensuring that only those users and entities with proper permissions can access data. **Note:** Encryption could impact Open Data Lake performance.

Data in Transit

Qubole transmits data that travels over public networks using strong encryption. This includes data transmitted between Qubole clients and the service. Qubole supports the latest recommended secure cipher suites to encrypt all traffic in transit, including use of Transport Layer Security (TLS) v1.2 (sometimes referred to as SSL or HTTPS) for communication between customer browsers and clients, REST API endpoints and Qubole services. Qubole has enabled this by default.

Qubole also monitors the changing cryptographic landscape and upgrades the cipher suite choices as the landscape changes, while also balancing the need for compatibility with older browsers and endpoints.

Data at Rest

Customers can configure compute clusters to use encryption when copying data to process to slaves and reporting results to the master node through the Qubole UI. This feature is located the Control Panel inside of the *Clusters Configuration/Advanced Configuration* drop-down menu.

Qubole also supports AWS S3 Server-Side Encryption, when enabled. This ensures that when data is written to S3, the data will immediately be encrypted on disk. This is configured within the Amazon AWS Console for the customer's account and is granted to Qubole via the IAM role in AWS. Refer to the documentation **here**.

Data Privacy and Data Integrity

Qubole ACID and Apache Ranger address distinct requirements of granular data access control and granular delete/merge/update respectively. Qubole supports ACID and Apache Ranger across multiple open source data processing engines, such as Apache Spark, Presto, or Hive and multiple clouds. Enterprises can now implement a centralized policy definition and management across engines and clouds without required technical expertise for each engine and public cloud provider. With Qubole's role-based access control (RBAC) integrations with Active Directory, LDAP, SAML2.0, organizations can leverage their existing RBAC solutions to manage user access to data lakes. Regarding data integrity supported via Qubole ACID, users can make inserts, updates, and deletes on transactional Hive Tables—defined over files in a data lake via Apache Hive—and query the same via Apache Spark or Presto. Qubole ACID also supports writes via Spark. Qubole writes only to changed rows, thus providing faster rewrites, updates, and deletes. It is most efficient and performant with built-in write and read optimizations

PLATFORM SECURITY

Firewalling via Security Groups

Qubole applies strict security measures when creating and maintaining customer clusters. Qubole dynamically creates security groups around clusters and creates access keys to clusters and dynamically encrypts ephemeral storage with one-time use keys that are destroyed on cluster terminate and are never stored. This is enforced (at a minimum) with secure permissions granting access from the Qubole Orchestration Tier to the Data Plane which resides in the customer's AWS account. All communication between the Orchestration Plane and the Data Plane is encrypted.

Security Event and Incident Management

Qubole maintains a Security Event and Incident Management System (SEIM) which is a central platform where alerts from a variety of sources are forwarded for review. In the event an event exceeds a given threshold, the event is forwarded to an on-call engineer for evaluation and escalation.

Intrusion Detection

Qubole deploys a log and kernel-based intrusion detection system that captures system changes, anomalous behaviour, and identifies anomalies and spurious behaviour on instances within the Qubole Orchestration Tier. When there is an unusual event detected in system or audit logs or in commands issued or series of unusual events, alarms are triggered in accordance with industry best practices and are forwarded to the SEIM or to an on-call engineer. Qubole supports customer-supplied and operated Intrusion Detection (IDS) or Intrusion Prevention (IPS) services within the customer account via bootstrapping. Your solutions architect can explain how this works and you can find additional documentation **here**.

File Integrity

Qubole operates a file integrity (FIM) management service that detects changes and acts as a defence against system compromise, malicious behaviour, and unauthorized actions by monitoring certain critical files on instances within the Qubole Orchestration Tier such as changes to critical logs, Qubole application code, SetUID binaries and additional configuration files and objects that may indicate a host has had a configuration change that should be investigated. When an unusual file event is detected, an alert is forwarded to the Qubole SEIM.

Vulnerability Scanning and Detection

Qubole externally and internally scans systems to determine if there are any application or system-specific risks. Qubole has a dedicated operations team who work around-the-clock to remediate these risks. Vulnerability detection is used to identify known software vulnerabilities in the packages, applications, operating systems, databases, and services used in Qubole Open Data Lake Platform. These scans are configured to run weekly and are remediated based on severity.

Hardened Baselines and Configuration Standards

Qubole evaluates the baseline of systems deployed in the Qubole Orchestration Tier using the Center for Internet Security (CIS) benchmarks. These benchmarks cover not only applications installed, configured, and running on systems, but also system and configuration files including restrictions around permissions, logs, system binaries. This benchmark covers more than 150 discrete changes and is applied to all instances deployed to both the Qubole Orchestration Tier as well as instances launched in customer accounts via the Qubole Data Plane Amazon Machine Instance (AMI).

Customer-Supplied Keys

Qubole also supports Amazon AWS Key Management Service (KMS). This service stores keys in a shared Hardware Security Module (HSM) dedicated to encryption key storage. For more information on how to configure and use this feature, please refer to the AWS documentation.

Authentication

To further reduce the risk of unauthorized data access, Qubole employs multi-factor authentication for all administrative access to systems with more highly sensitive and regulated data. Where possible and appropriate, Qubole uses unique private keys for authentication. For example, at this time, administrative access to production servers requires operators to connect using both SSH and a one-time password associated with a device-specific token.

Where passwords are used, multi-factor authentication is enabled for access to sensitive and regulated data. Passwords are required to be complex. System passwords are auto-generated to ensure uniqueness. These system passwords and accounts must have passwords longer than 20 characters and cannot consist of a single dictionary word (among other requirements). Qubole requires personnel to use an approved password manager. Password managers generate, store and enter unique and complex passwords. Using a password manager helps avoid password re-use, phishing, and other behaviours that can reduce security.

System Monitoring, Logging, And Alerts

Qubole monitors servers and workstations to retain and analyse a comprehensive view of the security state of its corporate and production infrastructure. Administrative access, use of privileged commands, and system calls on all servers in Qubole's production network are logged. Qubole's Security Team collects and stores production logs for analysis. Logs are stored in a separate network. Access to this network is restricted to members of the Security Team. Logs are protected from modification and retained for two years. Log analysis is automated to the extent practical to detect potential issues and alert responsible personnel. Alerts are examined and resolved based on documented priority.

We take a variety of steps to combat the introduction of malicious or erroneous code to our operating environment and protect against unauthorized access.

Controlling Change

To minimize the risk of data exposure, Qubole controls changes, especially changes to production systems, very carefully. Qubole applies change control requirements to systems that store data at higher levels of sensitivity. These requirements are designed to ensure that changes potentially impacting Customer Data are documented, tested, and approved before deployment.

Preventing and Detecting Malicious Code

In addition to general change control procedures that apply to our systems, Qubole production network is subject to additional safeguards against malware.

Server Hardening

New servers deployed to production are hardened by disabling unneeded and potentially insecure services, removing default passwords, and applying Qubole's custom configuration settings to each server before use.

Disaster Recovery and Business Continuity

Qubole uses services provided by its cloud provider to distribute its production operation across multiple physical locations. These locations are within one geographic region, but protect Qubole service from loss of connectivity, power infrastructure, and other common location-specific failures. Production transactions are replicated among these discrete operating environments, to protect the availability of Qubole service in the event of a location-specific catastrophic event. Qubole also retains a full backup copy of production data in a remote location more than 2500 mile from the location of the primary operating environment. Full backups are saved to this remote location once per day and transactions are saved continuously. Qubole tests backups at least quarterly to ensure they can be correctly restored.

3rd-Party Suppliers

To run its business efficiently, Qubole relies on sub-service organizations. Where those sub-service organizations may impact the security of Qubole production environment, Qubole takes appropriate steps to ensure its security posture is maintained. Qubole establishes agreements that require service organizations adhere to confidentiality commitments Qubole has made to its users. Qubole monitors the effective operation of the organization's safeguards by conducting reviews of its service organization controls before use and at least annually.

QUBOLE ON AMAZON AWS

Natively designed for AWS and tightly integrated with its storage, compute, and other key architectural elements, running Qubole - Open Data Lake Platform on AWS provides unparalleled enterprise-level security and data governance. Account provisioning is straightforward and, unlike other big data as-a-service providers, Qubole does not charge you to setup enterprise authentication and authorization that supports all your business needs.

Defining Your Security Options in AWS

Setting up a Qubole account and leveraging AWS security features is a simple process and provides virtually unlimited flexibility. Qubole supports SAML, OAuth, ADFS, Whitelisting and granular access controls and, while it is your responsibility to create and define your accounts, grant access to data, and configure your cloud services, Qubole support is always available to assist you.

Hive Authorization in AWS—Improving Enterprise-level Security and Data Governance in the Cloud

Qubole has defined a new security model integrating AWS storage authorization with Hive authorization. This improves usability for both cloud-storage administrators and data administrators, while eliminating errors that arise from end-user authorization problems. This is an important milestone on the way to Qubole's goal of building a secure, enterprise-level cloud platform. Qubole is one of the first vendors to add cloud storage-level checks at query compile time, and consequently offers the most secure platform in the cloud. For more information about Qubole Open Data Lake Platform and Hive, please see: **Security Model Authorization**

Creating an Account and Authorizing Users

To create a Qubole account, visit **Qubole** and click the sign-up link in the lower left-hand corner. For security reasons, you will be sent an email to confirm and activate your account.

Single Sign-On

Enterprise Ready

 **OAuth 2.0**

 **SAML v2**

 **ADFS**

Qubole supports all industry-standard authentication and authorization protocols including OAuth 2.0, SAML 2.0 and ADFS. You can also use the Qubole user management features to administer your Qubole users. To locate this feature in Qubole, navigate to *Control Panel* > *Manage Users.*

OAuth

Open Authentication (OAuth) can be configured for use with Qubole. OAuth extends a token validated against another form of authorization like Google. Qubole supports Google OAuth 2.0, Microsoft OAuth 2.0. To use OAuth please find a reference **here** for more detailed setup instructions.

SAML

Security Assertion Markup Language (SAML) is an enterprise authentication and authorization schema that provides for centralized authentication and provisioning of users from a secondary source that can be controlled by the enterprise. SAML is an open, XML-based standard that can be configured using several third parties including Okta, OneLogin, Ping Identity and others who support SAML v2.0.

Active Directory Federated Service (ADFS)

ADFS allows an enterprise to connect to a corporate Active Directory to enable authentication that aligns with existing corporate needs.

Accessing Data Securely

Qubole takes advantage of Identity and Access Management (IAM) roles in AWS to limit access to resources such as storage and compute. Using a refined set of permissions allows our customers to use Qubole on their behalf. Qubole also provides the right permissions for your data analysts and operators to create, run and modify queries and commands. Qubole's extensive functionality allows you to configure 14 different types of resources. Additionally, common concerns are addressed including limiting access rights to modify or affect the status of clusters, limiting the types of commands your users can execute and the data engines they can use. For more information, see **Managing Roles in Qubole Open Data Lake Platform.**

Manage Roles

Role Name	Policy			Action
	Access	Resource	Policy Actions	
dashboard-user	allow	Clusters	start	
	allow	Notebook Dashboards	read, execute	
	allow	Folder	read	
system-admin	allow	All	all	
system-user	allow	All	read	
	allow	Commands	create	
	allow	Clusters	start	
	allow	Templates	create, clone, run	
	allow	Workspace	create, update	
	allow	Account	auth_token	
	allow	Scheduler	create, clone	
	allow	Scheduler Instance	kill, rerun	
	allow	App	create	

How IAM Roles Work To Manage Secure Access And Authorization

Qubole uses AWS IAM roles to limit the privileges required to use Qubole Services. Specifically, IAM roles allow Qubole customers to limit privileges to only allow initiating and terminating instances.

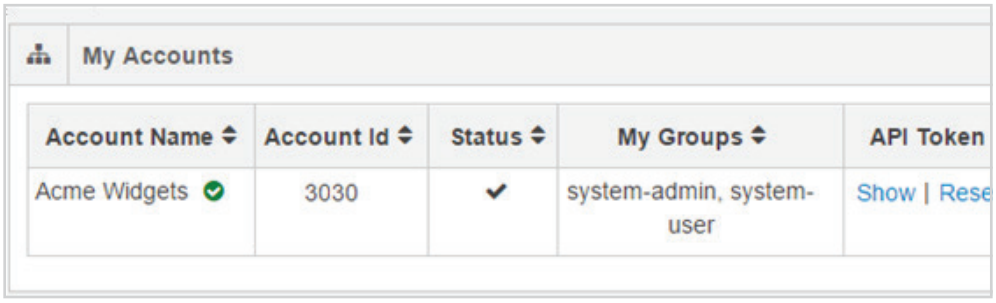
Those instances are then formed into clusters that you define, and you grant a secondary role that allows Qubole to attach a data storage policy to those resources. This secondary or (dual) role only operates within the customer's account and is not accessible outside of that account by anyone including Qubole.

Qubole will initiate and terminate instances on your behalf based on your configuration which you set inside of the Qubole Open Data Lake Platform Control Panel. This includes minimum and maximum cluster size, spot or on-demand, or a combination of both using fallback.

Per-User API Tokens

Each user can also be configured with their own API token allowing granular access controls and auditing on a per-user basis. This provides you with maximum control over your users when used in conjunction with roles described above. This token can also be reset to comply with more stringent security controls.

You can locate your (user) authentication token by navigating in Open Data Lake Platform to *Control Panel* > *My Accounts* > *Show API Token.*



CONCLUSION

We take security seriously at Qubole and we pride ourselves on the vigilance we employ to protect our customers' data assets. We believe that a mature security organization requires coordinated dedication across technology, policy, procedures, and people. This dedication is underscored by the risk-based approach laid out in this document to demonstrate strength at every layer of security. From compliance audits to HIVE authorization in AWS, our security strategy is designed to minimize any potential vulnerability or weakness. We want our customers to know their data is sufficiently protected by this approach and welcome the opportunity to discuss these practices further.

Learn More

For the latest information about our product and services, please see the following resources:

Qubole Whitepapers

Qubole Case Studies

Qubole Technical Documentation

You can visit the AWS Marketplace anytime to get up and running with Qubole!

TRY QUBOLE IN AWS TODAY!

For more information:

Contact:
sales@qubole.com

Try Qubole Open Data Lake Platform for Free:
AWS Marketplace: Qubole Open Data Lake Platform

469 El Camino Real, Suite 205
Santa Clara, CA 95050
(855) 423-6674 | info@qubole.com

WWW.QUBOLE.COM